ConQuest 1.12
User Guide and Tutorials

2010 CSDS Release

Copyright © 2009 The Cambridge Crystallographic Data Centre
Registered Charity No 800579

Conditions of Use

The Cambridge Structural Database System (CSD System) comprising all or some of the following: ConQuest, Quest, PreQuest, Mercury, (Mercury CSD and Materials module of Mercury), VISTA, Mogul, IsoStar, SuperStar, web accessible CSD tools and services, WebCSD, CSD Java sketcher, CSD data file, CSD-UNITY, CSD-MDL, CSD-SDfile, CSD data updates, sub files derived from the foregoing data files, documentation and command procedures (each individually a Component) is a database and copyright work belonging to the Cambridge Crystallographic Data Centre (CCDC) and its licensors and all rights are protected. Use of the CSD System is permitted solely in accordance with a valid Licence of Access Agreement and all Components included are proprietary. When a Component is supplied independently of the CSD System its use is subject to the conditions of the separate licence. All persons accessing the CSD System or its Components should make themselves aware of the conditions contained in the Licence of Access Agreement or the relevant licence.

In particular:

- The CSD System and its Components are licensed subject to a time limit for use by a specified organisation at a specified location.
- The CSD System and its Components are to be treated as confidential and may NOT be disclosed or re-distributed in any form, in whole or in part, to any third party.
- Software or data derived from or developed using the CSD System may not be distributed without prior written approval of the CCDC. Such prior approval is also needed for joint projects between academic and for-profit organisations involving use of the CSD System.
- The CSD System and its Components may be used for scientific research, including the design of novel compounds. Results may be published in the scientific literature, but each such publication must include an appropriate citation as indicated in the Schedule to the Licence of Access Agreement and on the CCDC website.
- No representations, warranties, or liabilities are expressed or implied in the supply of the CSD System or its Components by CCDC, its servants or agents, except where such exclusion or limitation is prohibited, void or unenforceable under governing law.

Licences may be obtained from:

Cambridge Crystallographic Data Centre
12 Union Road
Cambridge CB2 1EZ
United Kingdom

Web: *http://www.ccdc.cam.ac.uk*
Telephone: +44-1223-336408
Email: *admin@ccdc.cam.ac.uk*

(UNITY is a product of Tripos, L.P. and MDL is a registered trademark of Elsevier MDL)

1 THE CAMBRIDGE STRUCTURAL DATABASE (CSD) (see page 6)
2 GENERAL FEATURES OF CONQUEST (see page 13)
3 SUBSTRUCTURE SEARCHES (see page 47)
4 TABULATING GEOMETRIES; 3D SUBSTRUCTURE SEARCHES (see page 79)
5 NONBONDED CONTACT SEARCHES (see page 96)
6 PEPTIDE SEARCHES (see page 108)
7 AUTHOR AND JOURNAL SEARCHES (see page 113)
8 COMPOUND NAME and Chemical class SEARCHES (see page 116)
9 CHEMICAL ELEMENT SEARCHES (see page 120)
10 CHEMICAL FORMULA SEARCHES (see page 123)
11 SPACE GROUP AND CRYSTAL SYSTEM SEARCHES (see page 126)
12 UNIT CELL SEARCHES (see page 129)
13 CRYSTAL DENSITY AND UNIT-CELL CONTENT SEARCHES (see page 131)
14 SEARCHES FOR EXPERIMENTAL DETAILS (see page 133)
15 GENERAL TEXT SEARCHES (see page 135)
16 REFCODE SEARCHES (see page 140)
17 FILTERS: SECONDARY SEARCH CRITERIA (see page 143)
18 VIEWING and SEARCHING IN-HOUSE DATABASES, DATABASE SUBSETS and Previous CSD Versions (see page 145)
19 USING CONQUEST FOR TEACHING - CLASSROOM CONQUEST (see page 151)
20 RUNNING ConQuest IN BATCH MODE (see page 153)

21 TUTORIALS (see page 164)

22 APPENDIX A: GLOSSARY AND TECHNICAL DETAILS (see page 208)
23 APPENDIX B: CHANGES IN JOURNAL NAMES (see page 248)
24 APPENDIX C: ACKNOWLEDGMENTS (see page 250)

1 THE CAMBRIDGE STRUCTURAL DATABASE (CSD)

1.1 Chemical Coverage of the CSD

The CSD contains X-ray and neutron diffraction analyses of carbon-containing molecules having up to 1000 atoms (including hydrogens):

- Organics
- Compounds of the main group elements
- Organometallics
- Metal complexes

The CSD covers peptides of up to 24 residues; higher oligomers are covered by the Protein Data Bank (see page 237).

The CSD covers mono-, di- and tri-nucleotides; higher oligomers are covered by the Nucleic Acids Database (see page 234).

There is a small overlap between the CSD and the Inorganic Crystal Structure Database (ICSD) (see page 229) in the area of molecular inorganics. Information on other metal-containing structures can also be found in the Metals and Alloys Crystallographic Database (CRYSTMET) (see page 232).

1.2 Literature Coverage of the CSD

The CSD reflects the published literature. It is fully retrospective and is updated on a current basis. All information is abstracted from primary journals and associated supplementary information in both hard-copy and electronic forms. The CSD is the official depository for crystal structure data for more than 80 journals. The CSD also accepts direct deposition of crystal structure data, preferably electronically in CIF Format (see page 218). These are included in the database as Private Communications (see page 237).

1.3 Database Entries

Each entry in the CSD relates to an individual publication of a specific crystal structure. A paper that reports two or more crystal structure determinations will generate two or more entries in the CSD. Multiple determinations of the same structure (by the same or different authors) are included in the database. These may report improved structural models, or new experiments performed at different temperatures or pressures, or using different radiations.

Each CSD entry contains the following information (fully described elsewhere):

- Bibliographic, textual and numerical information
- Chemical diagram
- Crystal structure data
- Filters (secondary search criteria)

1.4 Refcodes (CSD Entry Identifiers) and Deposition Numbers

Each CSD entry is identified by a unique Refcode (entry ID) (see page 238) comprising:

- Six letters, e.g. ABACOF
- Two digits identifying additional structure determinations, e.g. ABACOF03

The CSD also records a variety of Deposition Numbers (see page 224) published by many primary journals to identify crystallographic data deposited with the CCDC.

1.5 Information Content of the CSD

The searchable information fields for a CSD entry may include those listed above.

1.5.1 Bibliographic Information in the CSD

- Authors' Names (see page 213)
- Journal Name (see page 229)
- Journal Volume (see page 230)
- Journal Page (see page 230)
- Journal Year
- Publication DOI

Further information is also stored with the above information, e.g. cross references to related structures, details of re-interpretations or re-determinations, or if the entry is a has come from a source other than a the published literature (see Private Communications, page 237).

Approximately three quarters of the structures in the database also have a hyperlink to the article that discloses the structure. This is normally accessed via CrossRef (*http://www.crossref.org*) though links are also available where appropriate through the IUCr. A subscription to the publication in question will normally be required to view the article.

1.5.2 Chemical Information in the CSD

- Compound name (see Compound Name Conventions, page 219)
- Chemical formula (used for element and formula searches) (see Chemical Formula, page 218)
- Chemical diagram (used for 2D and 3D substructure searches, non-bonded interaction searches) (see Chemical Diagram, page 217)

Further chemical property information is also available, e.g. absolute configuration, bioactivity, conformer, isomer, racemate, where provided in the published literature.

1.5.3 Crystallographic Information in the CSD

- Atomic Coordinates (see page 212)
- Unit Cell Parameters (see page 242)
- Reduced Cell Parameters (see page 238)
- Crystal Systems (see page 222)
- Space Group (see page 240)
- Crystal density, colour, habit and source of crystal
- Number of formula units per unit cell (Z) (see Z (Number of Molecules per Unit Cell), page 246)
- Number of formula units per asymmetric unit (Z') (see Z' (Number of Molecules per Asymmetric Unit), page 247)
- R-factor (see page 239)

- Average e.s.d. of C-C Bonds (see page 214)
- Temperature of Structure Determination (see page 241)
- Disordered structure flag (see Disordered Structures, page 224)
- Unresolved errors flag (see Unresolved Errors, page 243)

1.5.4 Text Information in the CSD

- Details of disorder (see Disordered Structures, page 224)
- Details of corrected or residual errors
- Details of non-standard experimental conditions
- Special text strings: drug activity, polymorphism, absolute configuration, etc.
- General remark

1.5.5 Filters: Secondary Search Criteria

The CSD contains a number of flags indicating the presence or absence of certain characteristics in each entry. These flags may be used as secondary filters to restrict the output of searches to, e.g., error-free, non-disordered structures.

- Atoms with 3D Coordinates (see page 234)
- R-factor within preset range (see page 239)
- Not disordered (see Disordered Structures, page 224)
- No errors (see Unresolved Errors, page 243)
- Not polymeric (see Polymeric Bonds, page 236)
- No ions
- No powder structures (see Powder Data, page 237)
- Only Organics (see page 235)
- Only Organometallics (see page 235)

1.6 A Typical CSD Entry

Bibliographic and chemical text information:

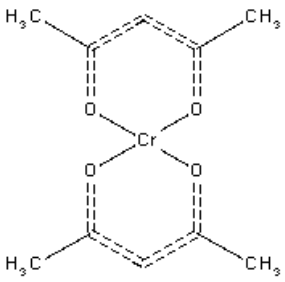
Formula: C10 H14 Cr1 O4
Name: bis(2,4-Pentanedionato) chromium

Chemical Diagram

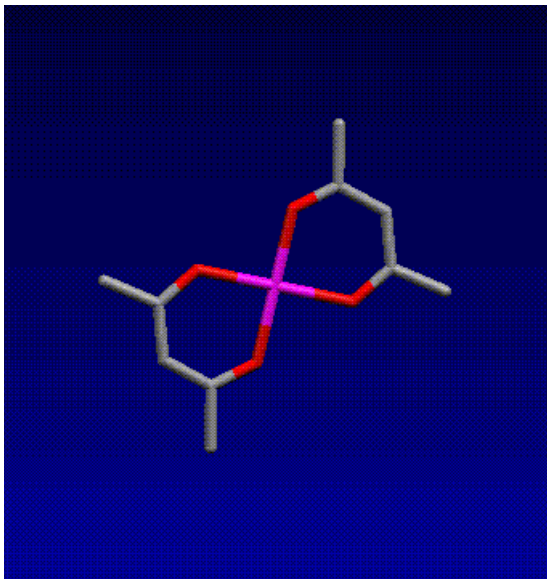
Spacegroup
Name: P21/n **Number:** 14

Cell Parameters
a: 11.445 **b:** 4.748 **c:** 10.325
alpha: 90.00 **beta:** 91.60 **gamma:** 90.00
Volume: 560.851

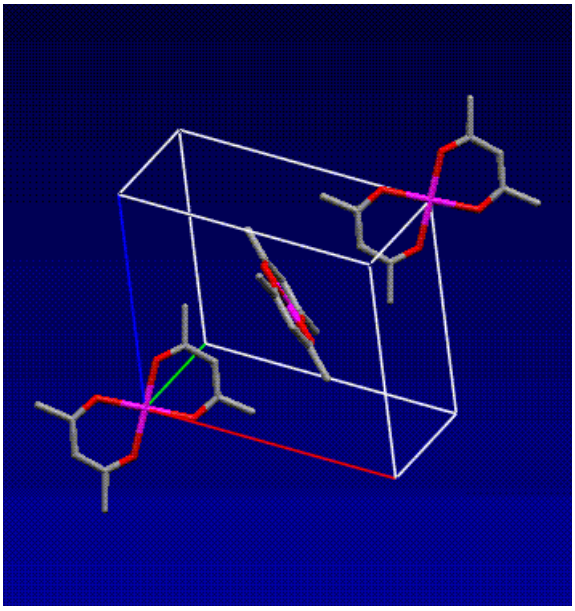
Reduced Cell Parameters
a: 4.748 **b:** 10.325 **c:** 11.445
alpha: 91.60 **beta:** 90.00 **gamma:** 90.00
Volume: 560.848



3D Molecular Structure



3D extended crystal structure (crystal packing):



- Not all of the possible CSD information fields are present in all entries
- Atoms of the 2D chemical diagram are mapped onto those of the 3D molecular structure to provide a complete representation of the chemical and geometrical structure. This representation is used for all 2D and 3D substructure searches and, in conjunction with the symmetry operators of the crystallographic space group, is also necessary when searching for intermolecular interactions.
- In cases where molecular symmetry coincides with crystallographic symmetry, the CSD stores atomic coordinates for complete chemical molecules. Thus, although the crystallographer reported only the atomic coordinates for the Asymmetric Unit (see page 211) (one half molecule) in the example above, the CSD stores full coordinates for the symmetry-related atoms that complete the molecule. These atoms are identified by alphabetic suffixes in their atom labels. The 3D atomic coordinates in the CSD describe what is termed as the Crystal Chemical Unit (see page 221).

2 GENERAL FEATURES OF CONQUEST

2.1 The Basics of Using ConQuest

- Use the *Build Queries* screen (see Building Queries: Overview, page 14) to build one or more queries. Each query defines something to be searched for, e.g. a set of unit cell dimensions.
- If necessary, use *Combine Queries* (see Complex Query Combinations, page 21) to combine the queries using logical operations, e.g. to specify a search for structures containing arsenic but *not* antimony.
- Hit the **Search** button to start the search (see Starting Searches, page 27). This opens up the *Search Setup* dialogue box (see Naming Searches and Setting Filters, page 28), which can be used to set filters, e.g., for excluding structures with poor R-factors (see FILTERS: SECONDARY SEARCH CRITERIA, page 143).
- Use *View Results* (see Viewing Search Results, page 29) to look at the structures found by the search.
- Save the search results (see Saving and Reading Searches, page 41) or the queries (see Saving and Reading Queries, page 19), or run another search, or exit the program.
- New queries may be built and new searches started while an existing search is running. Use the tabs *Build Queries*, *Combine Queries* and *View Results* to move between different parts of the program.

2.2 Building Queries: Overview

- Queries are built in the *Build Queries* area. This is displayed by default when ConQuest is opened and is accessible at any time by hitting the tab labelled *Build Queries*.
- The buttons on the left of the *Build Queries* screen allow different types of queries to be built.

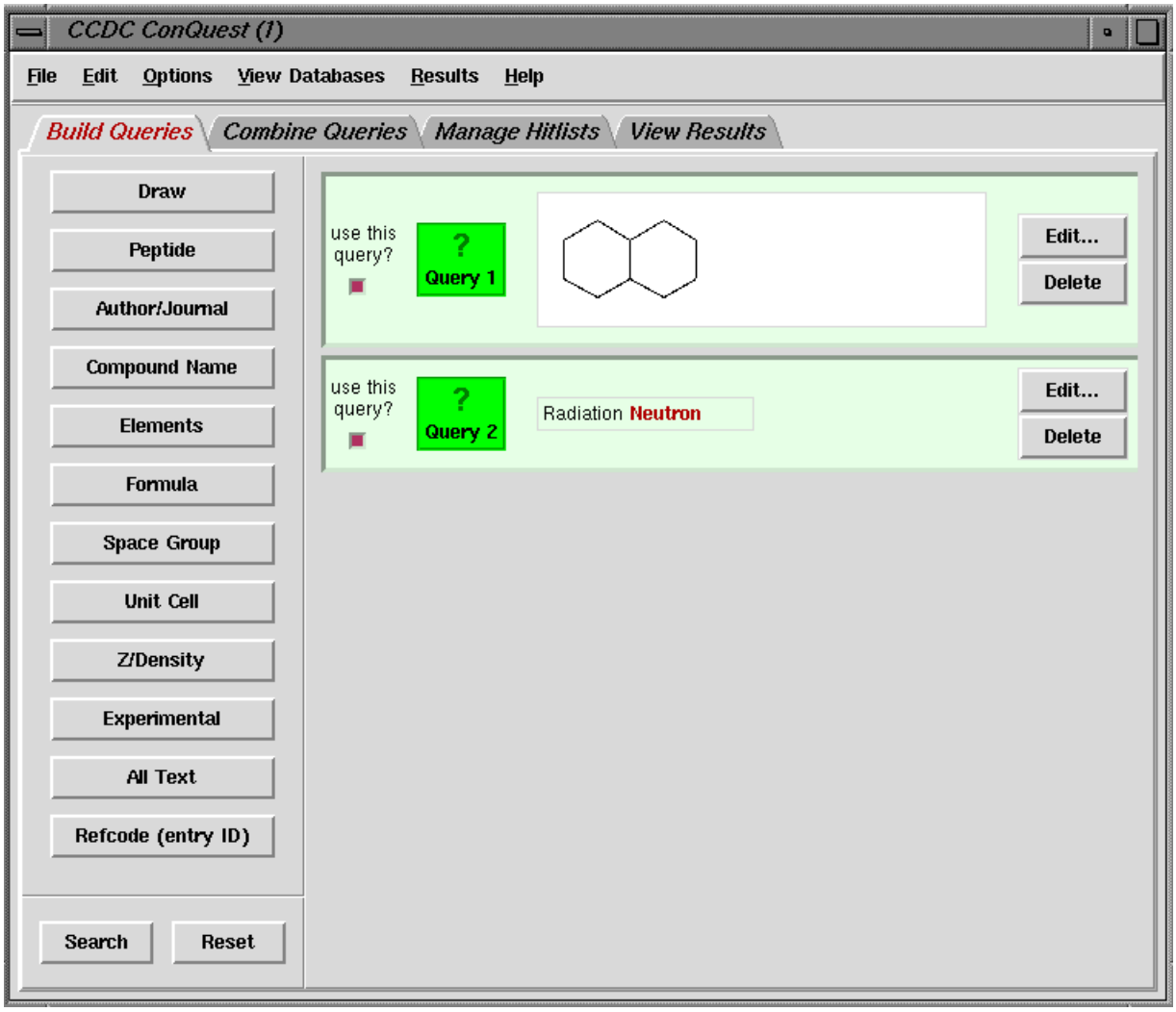
Button	***Type of Query***	***Example***
Draw (see SUBSTRUCTURE SEARCHES, page 47)	Substructure (2D or 3D)	Find structures containing a cyclopropane ring
Peptide (see PEPTIDE SEARCHES, page 108)	Peptide sequence	Find peptides containing the sequence: *ALA-PRO-PHE*
Author/Journal (see AUTHOR AND JOURNAL SEARCHES, page 113)	Bibliographic	Find structures published by *J.D.Dunitz* in *J.Am.Chem.Soc*.
Name/Class (see COMPOUND NAME and Chemical class SEARCHES, page 116)	Chemical name CSD Class	Find *morphine* Find all entries within the *Steroid* class
Elements (see CHEMICAL ELEMENT SEARCHES, page 120)	Chemical elements	Find structures containing *Au* and *Os*
Formula (see CHEMICAL FORMULA SEARCHES, page 123)	Chemical formula	Find molecules with the formula $C_6 H_{12} O_6$
Space Group (see SPACE GROUP AND CRYSTAL SYSTEM SEARCHES, page 126)	Crystallographic space group and crystal system	Find structures crystallising in space group *Fdd2*

Unit Cell (see UNIT CELL SEARCHES, page 129)	Cell dimensions	Find structures with *a=10.0, b=20.0, c=30.0, alpha = beta = gamma = 90.0*
Z/Density (see CRYSTAL DENSITY AND UNIT-CELL CONTENT SEARCHES, page 131)	Density and unit cell contents	Find structures with *Z = 16*
Experimental (see SEARCHES FOR EXPERIMENTAL DETAILS, page 133)	Experimental details	Find structures solved at low temperature
All Text (see GENERAL TEXT SEARCHES, page 135)	General text	Find entries containing the word *green*
Refcode (entry ID) (see REFCODE SEARCHES, page 140)	CSD refcode (entry identifier)	Find CSD entry *BAJCOE*

2.3 Storage of Queries During a ConQuest Session

- As queries are built, they are listed in the *Build Queries* area. In the example below, the user has built two queries:

  Query 1: a substructure search for decalin derivatives.
  Query 2: to find structures determined by neutron diffraction.



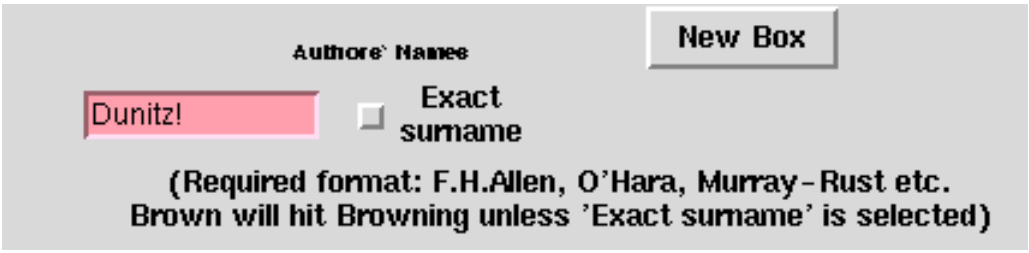
- Queries may be saved as files for later retrieval and re-use (see Saving and Reading Queries, page 19).

2.4 Use of Dialogue Boxes

Most ConQuest dialogue boxes have the same basic behaviour.

2.4.1 Typing into Input Boxes

- Boxes that can be typed into are coloured white or grey.
- Most input boxes turn pink if they contain invalid input. In the example below, the program has recognised that the character **!** is invalid in an author name.



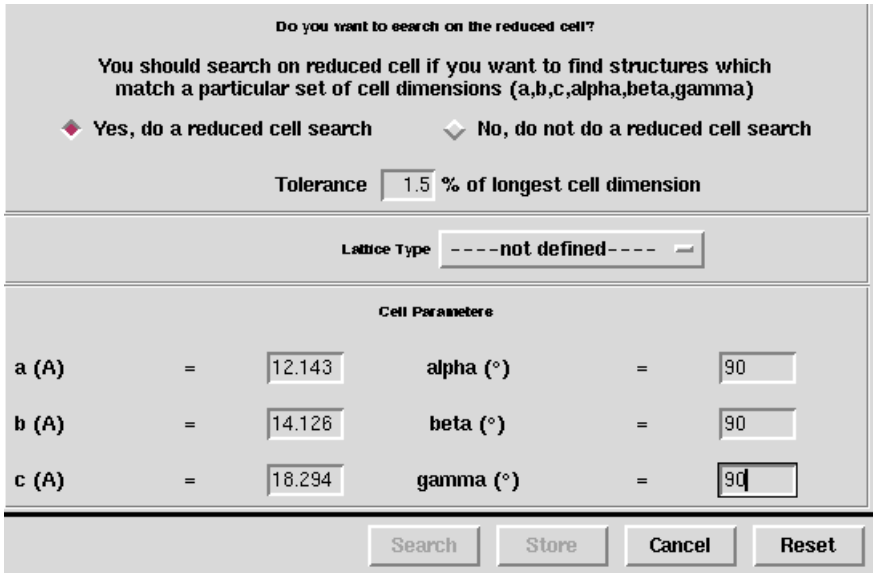
- Some dialogue boxes automatically correct or reject invalid input.

2.4.2 Case Sensitivity

Searches for text are always case insensitive, i.e. it does not matter whether upper or lower case is used.

2.4.3 Greyed-Out Buttons

- Buttons are greyed out when they cannot be used. This is usually because the user has not defined sufficient information for the button to be useable.
- In the example below, the **Search** button is greyed out because the user has not specified all the information required for a reduced cell search (specifically, the lattice type has not yet been defined).


Do you want to search on the reduced cell?
You should search on reduced cell if you want to find structures which match a particular set of cell dimensions (a,b,c,alpha,beta,gamma)
Yes, do a reduced cell search
No, do not do a reduced cell search
Tolerance 1.5 % of longest cell dimension
Lattice Type ----not defined----
Cell Parameters
a (A) = 12.143
alpha (°) = 90
b (A) = 14.126
beta (°) = 90
c (A) = 18.294
gamma (°) = 90
Search
Store
Cancel
Reset

2.4.4 Mathematical Operators

- Some queries involve testing the values of numbers, e.g. to find out whether a density is greater than a certain value. In such cases, the input box into which the number is typed is preceded by a button labelled with a mathematical operator such as = .

- Clicking on this button will produce a pull-down menu from which the required mathematical operator (=, not =, >, <, etc.) can be chosen.
- Using the = operator to test the value of a non-integral number can be unreliable, e.g. a density is unlikely to be *exactly* 2.000. Specify a **range** instead:

2.4.5 Closing and Resetting Dialogue Boxes

Most dialogue boxes have the following four buttons:

- **Search:** Starts a search with the query that has been defined in the dialogue box. Also closes the dialogue box and stores the query in the *Build Queries* and *Combine Queries* areas.
- **Store:** Stores the query in the *Build Queries* and *Combine Queries* areas and closes the dialogue box, but does not start a search.
- **Cancel:** Cancels all input and closes the dialogue box.
- **Reset:** Deletes all input but leaves the dialogue box open.

2.5 Keyboard Shortcuts

A set of keyboard shortcuts are available for some menu items. These are indicated next to the menu item, for example, in the top-level **File** menu:

2.6 Editing and Deleting Queries

- **Editing queries**: Go to *Build Queries* or *Combine Queries* and click on the **Edit...** button next to the query to be edited or double click on the contents of the query box.
- **Deleting queries**: Go to *Build Queries* or *Combine Queries* and click on the **Delete** button next to the query to be deleted. To undo this action, select **Edit** from the top-level menu bar and **Undelete Queries** from the resulting pull-down menu.

2.7 Saving and Reading Queries

- **Saving queries**: Go to *Build Queries* or *Combine Queries*, select **File** from the top-level menu bar and **Save Queries** from the resulting pull-down menu. It is then possible to save either one particular query or all queries (in a single file).
- **Reading queries**: Go to *Build Queries* or *Combine Queries*, select **File** from the top-level menu bar and **Read Queries...** from the resulting pull-down menu.
- MDL MOL files can be used as queries. If the MOL file contains geometric objects such as planes and centroids, these will normally be transferred to ConQuest. If the MOL file contains defined distances, angles and dihedral angles then these too will normally be transferred, so long as a corresponding geometric parameter can be defined in ConQuest. The following table indicates the equivalencies between MDL and ConQuest geometric features. Those MDL features for which no ConQuest feature exist are ignored by ConQuest. Names for features will be retained unless they are not present or are more than six characters long. In these cases names will be automatically assigned. If duplicate names are loaded the second incidence is renamed. Loading of Multi-MOL files will result in one query for each substructure loaded.

MDL ID	MDL Feature	CQ Feature	CQ Prefix
-1	Point defined by two points and a distance (in Angstroms)	Dummy point	DUM
-2	Point defined by two points and a percentage		
-3	Point defined by a point, a normal line, and a distance		
-4	Line defined by two or more points		
-5	Plane defined by three or more points	Plane	PLN
-6	Plane defined by a point and a line		
-7	Centroid defined by points	Centroid	CENT
-8	Normal line defined by a point and a plane		
-9	Distance defined by two points and a range	Distance	DIS
-10	Distance defined by a point, line, and a range		
-11	Distance defined by a point, plane, and a range	Distance	DPP
-12	Angle defined by three points and a range	Angle	ANG
-13	Angle defined by two lines and a range		
-14	Angle defined by two planes and a range	Angle	APP
-15	Dihedral angle defined by four points and a range	Torsion	TOR
-16	Exclusion sphere defined by a point and a distance		
-17	Fixed atoms		

- Queries saved from ConQuest are stored in files with the extension *.qry*. Queries which can be read into ConQuest have either a *.qry* or *.mol* extension.

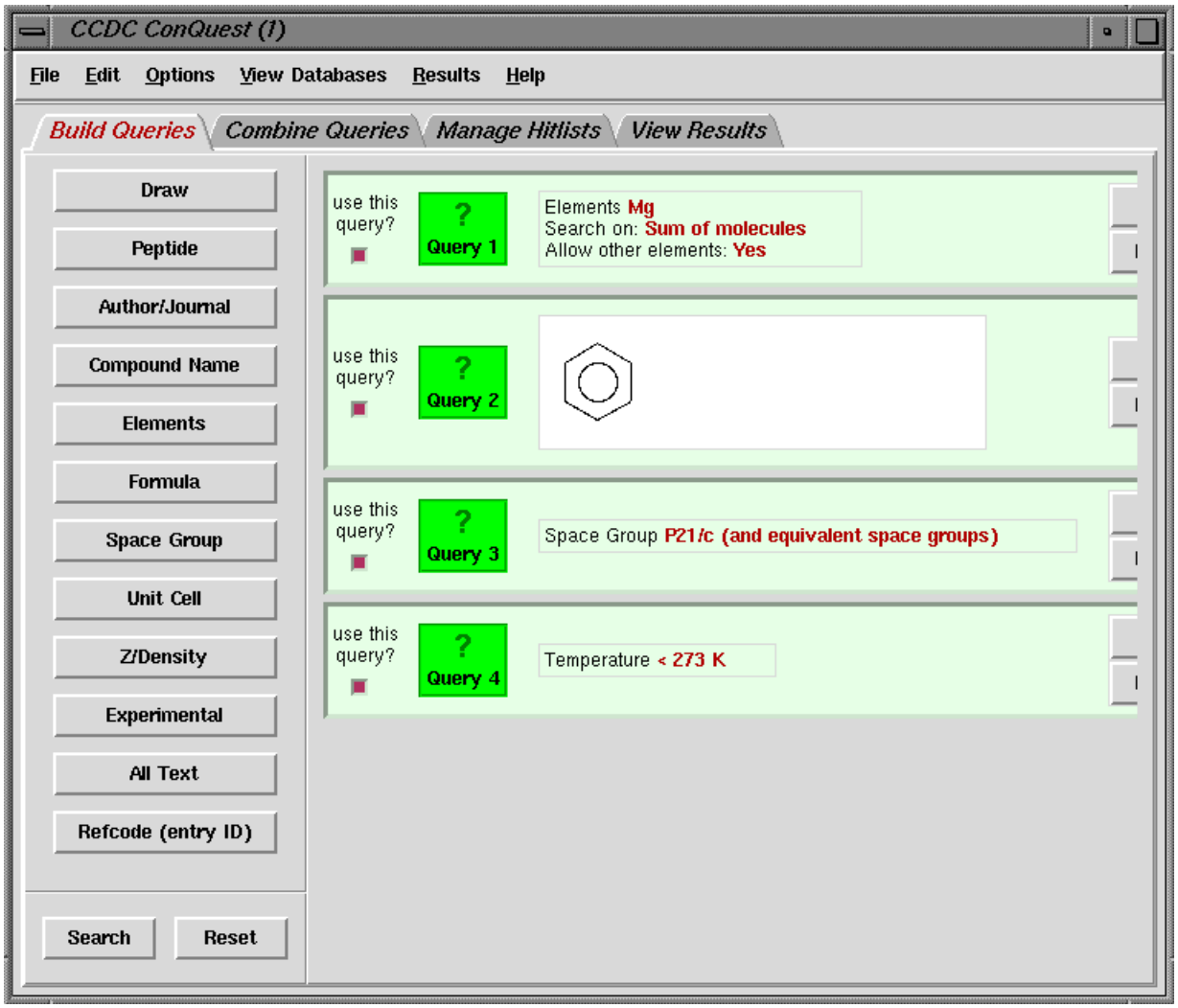
2.8 Combining Queries

2.8.1 Reasons for Combining Queries

- Finding entries that simultaneously satisfy two or more search conditions, e.g. contain arsenic *and* determined at low temperature.
- Finding entries that do not satisfy a search condition, e.g. do *not* contain arsenic.
- Finding entries that satisfy some search conditions but do not satisfy others, e.g. contain arsenic but were *not* determined at low temperature.
- Finding entries that satisfy at least one of a number of search conditions, e.g. contain arsenic *or* were determined at low temperature.

2.8.2 Simple Query Combinations

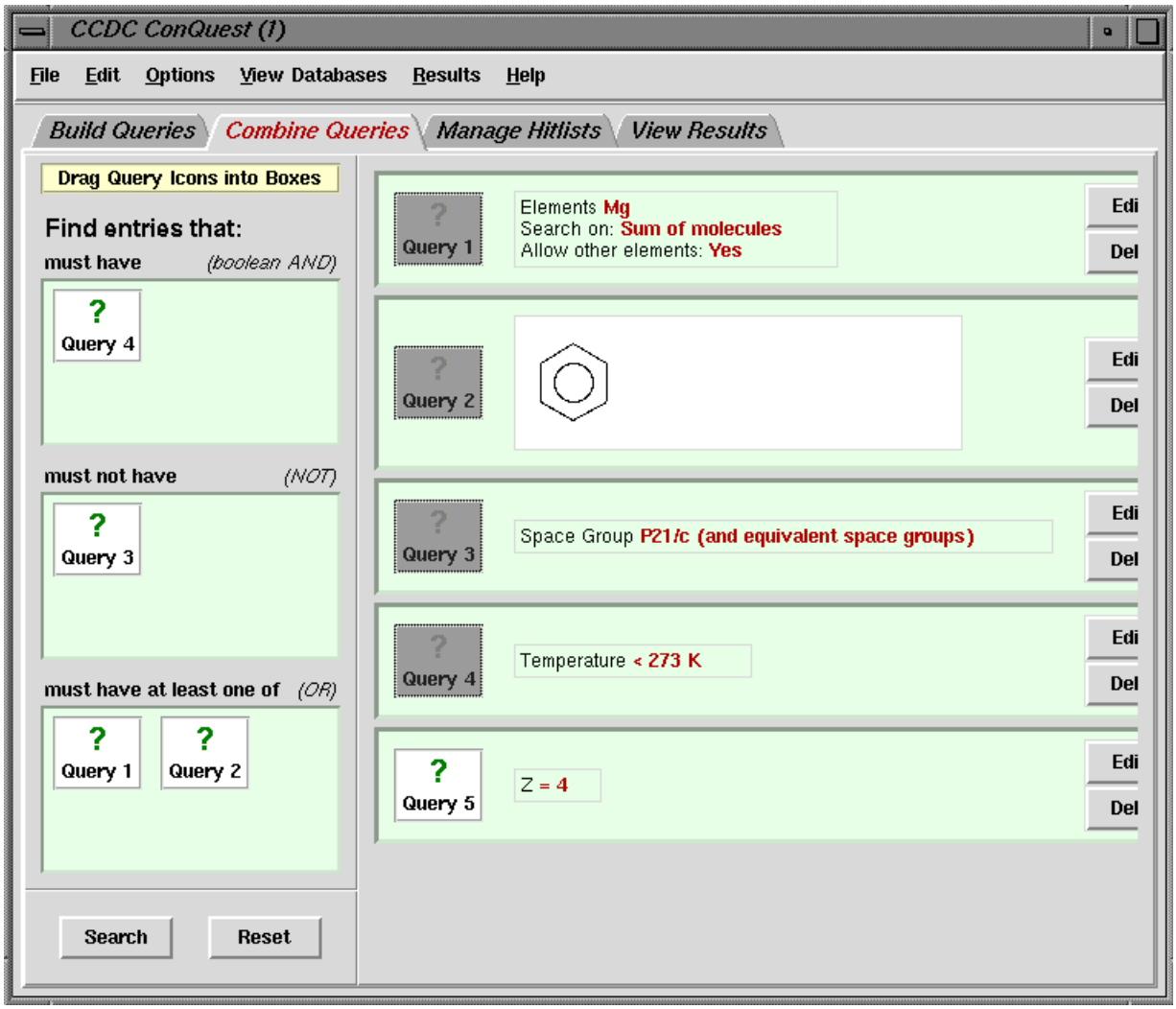
- Simple query combinations can be used to find structures which simultaneously satisfy two or more conditions, e.g. structures that contain magnesium *and* a benzene ring *and* were determined at a temperature below 273K.
- Build each query separately and store.
- Go to the *Build Queries* screen. Click on the check-boxes labelled **use this query?** (or on the query icons themselves) to turn each query *on* or *off*, as desired.
- Hit the **Search** button (bottom left-hand corner). The resulting search will find structures that satisfy all of the queries that are turned on.
- The example below will find structures that contain magnesium *and* a benzene ring *and* were determined at temperatures below 273K. Query 3 (space group P21/c) is turned off, so it is ignored.



2.8.3 Complex Query Combinations

- The *Combine Queries* screen can be used to find structures that satisfy a logical combination of queries, e.g. structures which contain magnesium *or* a benzene ring, do *not* crystallise in space group P21/c, *and* were determined at a temperature below 273K.
- Build each query separately and store (see Closing and Resetting Dialogue Boxes, page 18).
- Hit the tab labelled *Combine Queries*.
- The three boxes on the left-hand side of the *Combine Queries* screen (labelled *must have, must not have, must have at least one of*) represent the logical operations required to define complex searches. Drag query icons into the boxes to set up the desired search. The text above each of the logical operator boxes explains its function.
- An icon can be dragged by clicking on it with the left-hand mouse button and moving the mouse (keeping the button depressed) so that the icon *drops* into the desired box. Icons may also be dragged from box to box.
- Once the required search has been defined, hit **Search** (bottom left-hand corner) to start the search.

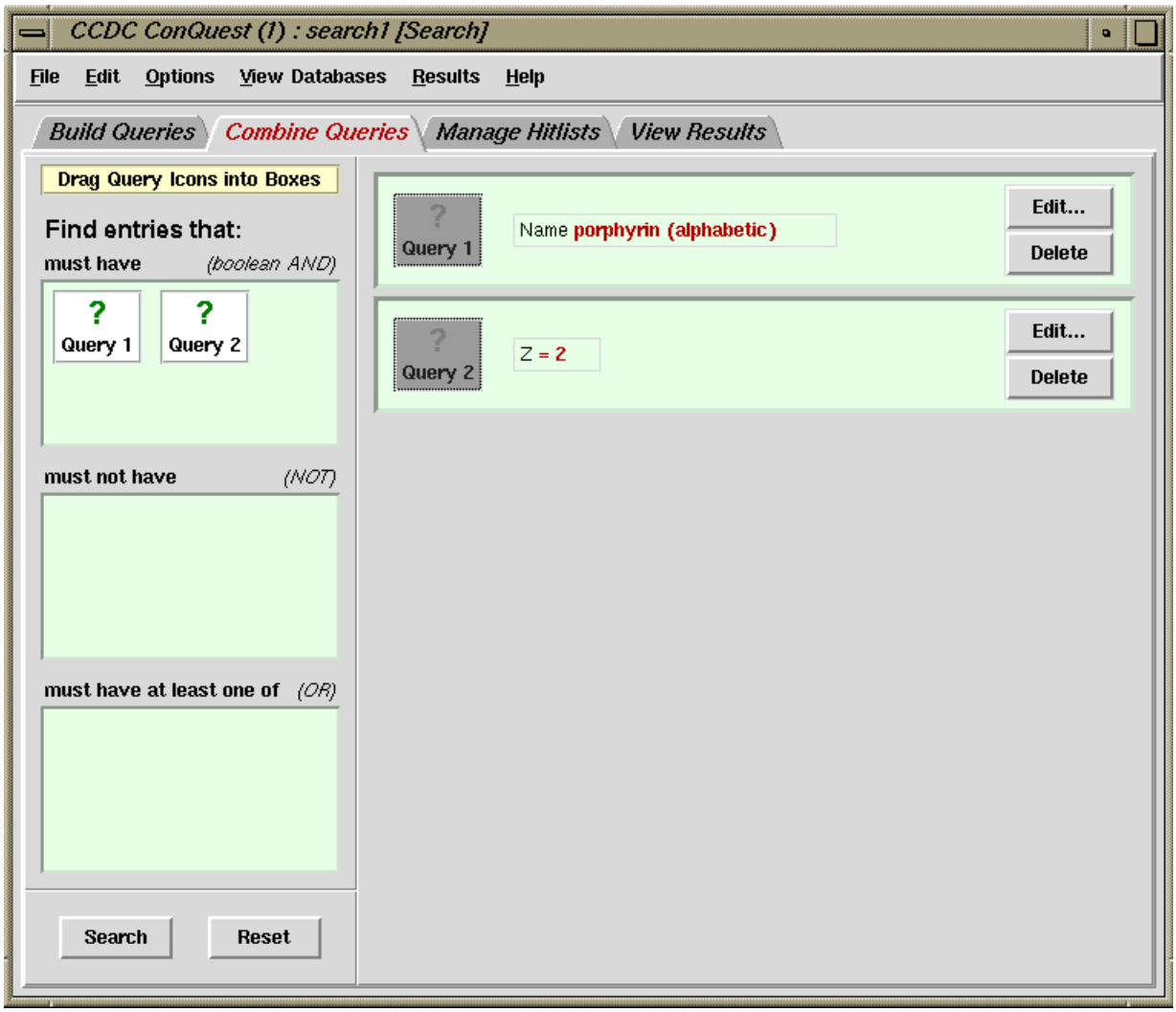
- The example below will find structures which contain magnesium *or* a benzene ring, do *not* crystallise in space group P21/c, *and* were determined at a temperature below 273K. Query 5, Z = 4, is ignored as it has not been dragged into a box.



2.9 Managing Hitlists

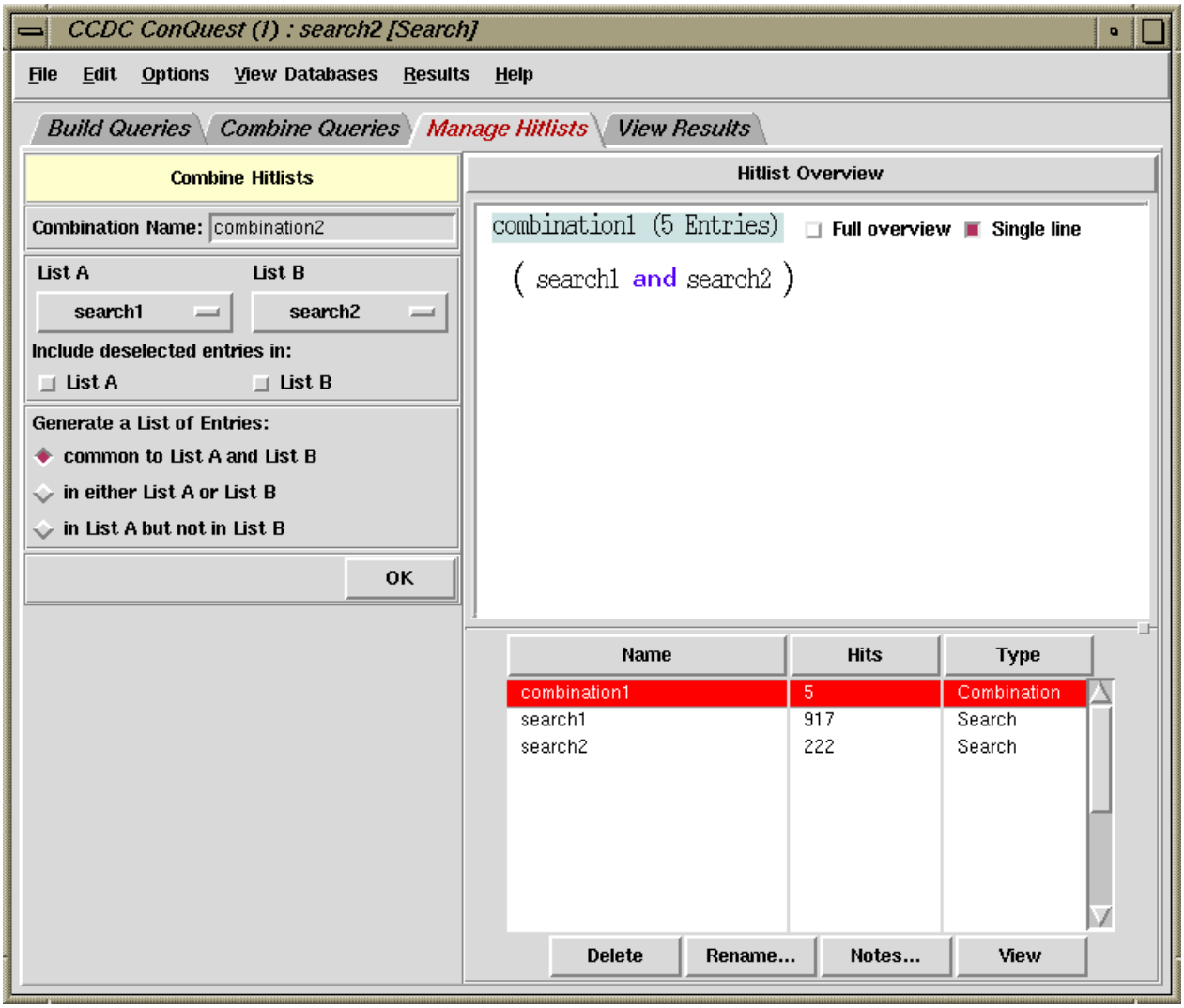
2.9.1 Hitlist Combinations

- A hitlist is a subset of CSD entries which can include search results, refcode lists or the results of combining these. The *Manage Hitlists* screen can be used to find CSD entries that satisfy a logical combination of hitlists, this can be a combination of search results, refcode lists or a combination of the results from either a search or a refcode list.
- Hitlist combinations are best illustrated with an example:
  - Set up the following search for all porphyrin structures having z=2 (see Simple Query Combinations, page 20):



  - Set up the search as described in Simple Query Combinations (see page 20) for all those structures that contain magnesium *and* a benzene ring *and* were determined at a temperature below 273K. You now have 2 sets of search results (i.e. hitlists) available which can be combined.

- Hit the tab labelled *Manage Hitlists*. On the left-hand side of the *Manage Hitlists* screen the pull-down menus under List A and List B have search1 listed as the default, select search2 from the pull-down menu under List B. Both the check-boxes next to List A and List B should be deselected (these can enable a combination of hitlists to include suppressed entries (see Suppressing Unwanted Entries, page 40).
- The three check-boxes listed on the left-hand side of the *Manage Hitlists* screen (labelled **common to List A and List B**, **in either List A or List B**, **in List A but not in List B**) represent the logical operations available for combining search results. Select **common to List A and List B** and hit **OK**.
- The results of combining these two hitlists, search1 and search2, are given in the *Hitlist Overview* pane and indicate that 5 entries are common to both searches. The default name for the hitlist combination, *combination1*, appears under *Name*, on the bottom right-hand side of the screen, along with the searches used to create this hitlist combination:



2.9.2 Saving and Reading Hitlist Combinations

- To save a combination of two hitlists, ensure that the hitlist combination is selected in the *Manage Hitlists* screen (if it is not, display it by using the **Results** top-level menu option, then select **Combinations** from the resulting pull-down menu to display the hitlist combination of your choice).
- Pick **File** from the top-level menu and **Save Hitlist Combination** or **Save Hitlist Combination As...** in the resulting pull-down menu**.** The first of these options will only be available if the hitlist has already been saved and since modified.
- Saved hitlist combinations will be given the extension *.chl* which can be read back into ConQuest and displayed in *View Results* as if the combination had just been created and modified.
- To read a previously saved hitlist combination, pick **File** from the top-level menu, followed by **Open** and then select **Hitlist Combination...** from the resulting pull-down menu. Loading a hitlist, either when reading from a file or when viewing hitlist combinations for the first time, may result in a progress window. If you have previously saved hitlist combinations, resulting from searches in the same ConQuest session, these can be viewed by selecting **Results** from the top-level menu followed by **Combinations** in the resulting pull-down menu to display the search of your choice.
- Alternatively, you can read in a hitlist combination, Refcode list or search, saved from an earlier ConQuest session, by picking **File** from the top-level menu followed by **Open** and then select **Recent Files** from the resulting pull-down menu.

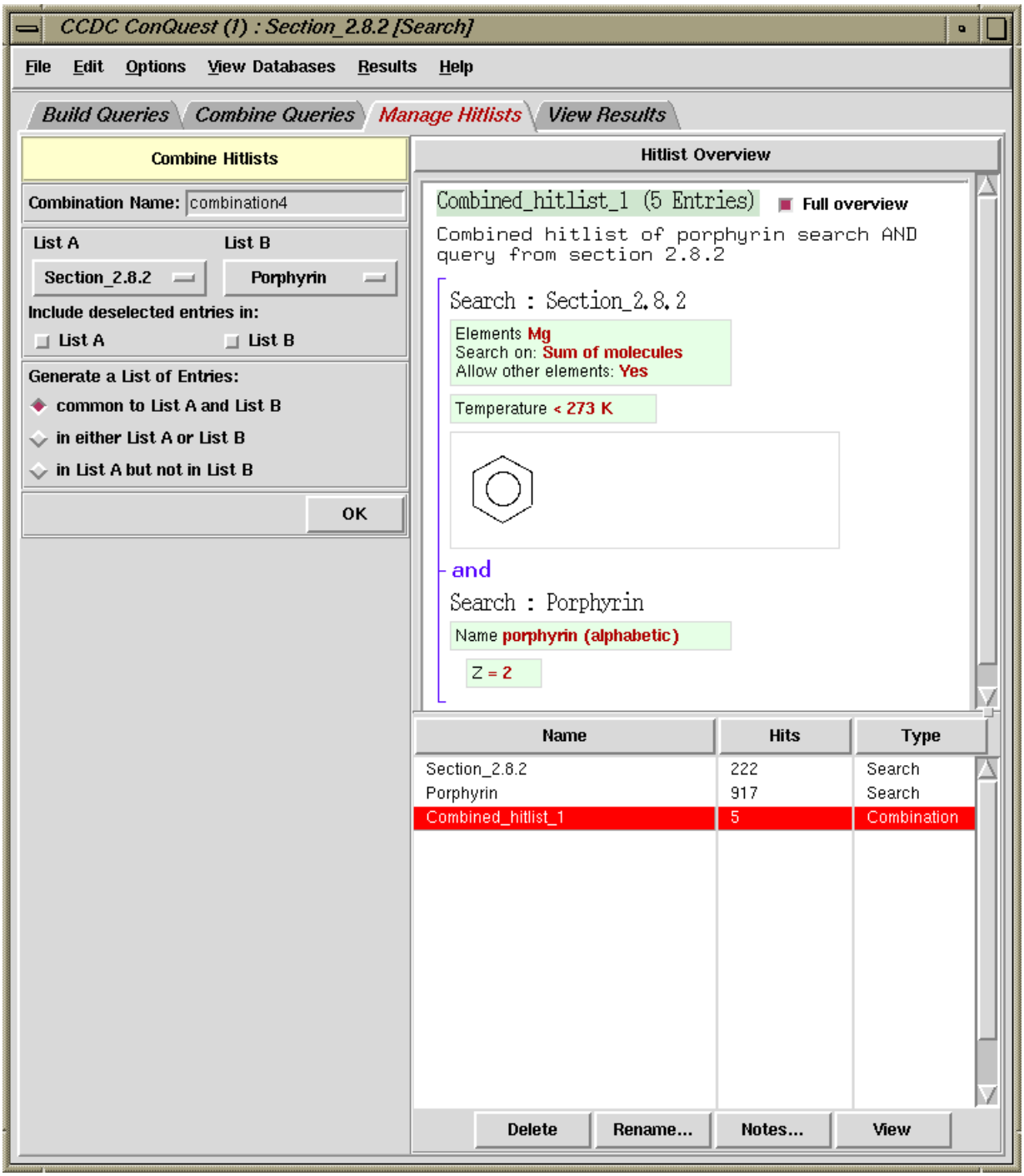
2.9.3 Renaming, Annotating, Viewing and Deleting Hitlists

Various options are available for renaming, annotating, viewing and deleting either searches or hitlists including combinations; some of these options also apply to Refcode lists:

- To rename a search or hitlist combination: Select it from list of hitlists (bottom right-hand side of the screen). Hit the **Rename...** button underneath this list. Type the new name into the *Edit hitlist name* input box. Hit **OK**. Alternatively, it is possible to provide a name for the combination in the *Combination Name* entry box on the top left-hand side of the screen. The name changes will appear immediately in the *Manage Hitlists* screen.
- To annotate a search or hitlist combination with additional information: Select it from the list of hitlists (bottom right-hand side of the screen). Hit the **Notes...** button underneath this list. Type any additional information into the *Hitlist Details* input box. Hit **OK** to include the additional information, **Reset** to clear any changes or **Cancel** to exit without making any changes.
- To view a search, hitlist combination or Refcode list from within the *Manage Hitlists* screen: Select it from list of hitlists (bottom right-hand side of the screen). Hit the **View** button underneath this list. You will be presented with the requested information in the *View Results* screen. Viewing a hitlist combination for the first time may result in a progress window. Alternatively, If you have previously created or loaded hitlists in the same ConQuest session

these can be viewed by selecting **Results** from the top-level menu followed by **Searches**, **Combinations** or **Refcode Lists** in the resulting pull-down menu to display the hitlist of your choice.

- The full overview for the hitlist combination is shown below; you can toggle between the full overview and the brief overview shown by default by selecting the appropriate check-boxes. In this example the hitlist combination has been annotated and renamed and both searches have been renamed to something more meaningful:



- To remove a search, Refcode list or hitlist combination from the current ConQuest session: Select it from the list of hitlists (bottom right-hand side of the screen). Hit the **Delete** button underneath this list, the change will appear immediately in the *Manage Hitlists* screen. If you wish to restore this information because, for example, it was deleted in error: Select **Edit** from the top-level menu, followed by **Restore Hitlists** and then select either **Searches**, **Combinations** or **Refcode Lists** from the resulting pull-down menu. Deleting hitlists in this way will *not* delete any saved files.

2.10 Starting Searches

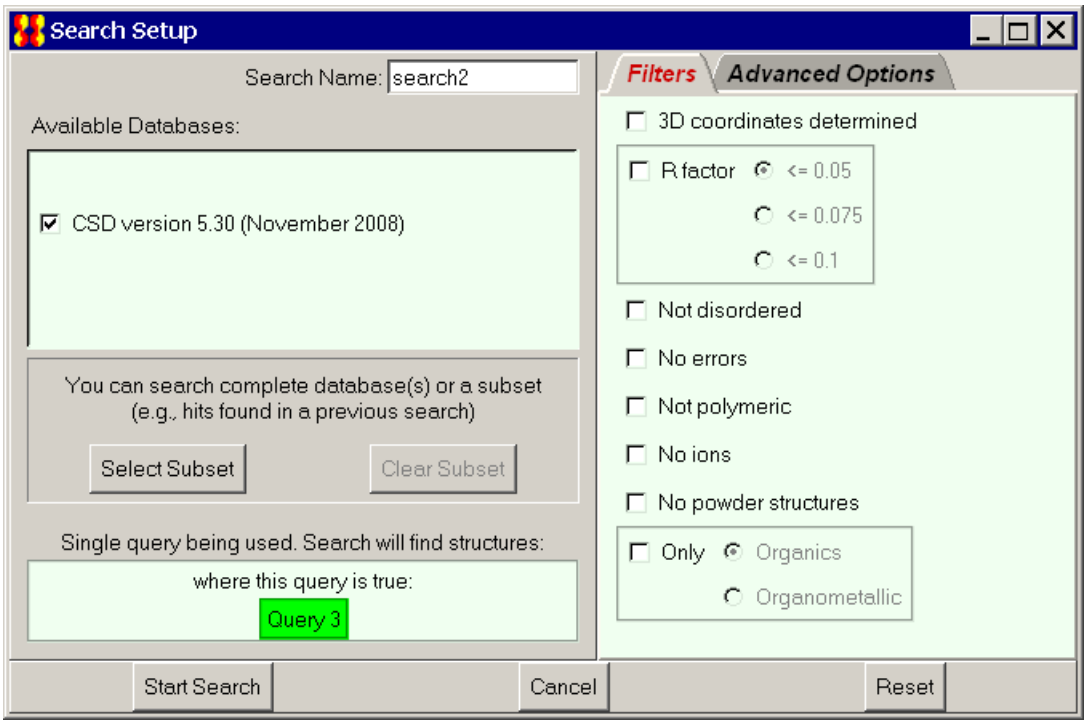
A search may be started by hitting a **Search** button. These can be found in several different places:

- Hitting **Search** in the *Draw* window (see Layout of the Draw Window, page 47) starts a search for the substructure query currently displayed in that window – *not* for any other query (including substructure queries) that may previously have been built and stored.
- Hitting **Search** in any of the dialogue boxes used to build queries, e.g. **Author/Journal**, starts a search for the current contents of that dialogue box – *not* for any other query that may previously have been built and stored.
- Hitting **Search** on the *Build Queries* screen starts a search for structures that satisfy all of the queries currently turned on, i.e. whose icons are coloured green (see Simple Query Combinations, page 20).
- Hitting **Search** on the *Combine Queries* screen starts a search for the logical combination of queries defined by the contents of the three *logical operator* boxes on the left-hand side (see Complex Query Combinations, page 21).

If a **Search** button is hit by accident, click on **Cancel** in the resulting *Search Setup* dialogue box.

2.11 Naming Searches and Setting Filters

Whenever a **Search** button is hit, the *Search Setup* dialogue box is opened before the search is actually started:



This box shows the version of the CSD that will be searched, summarises the query, and enables you to:

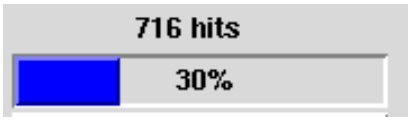
- Name the search by typing into the input box at the top left. This name is used throughout the ConQuest session, but a new name may be chosen if the search is saved as a file (see Saving and Reading Searches, page 41).
- Elect to run the search on a subset of the CSD, or on an in-house database of crystal structures, should you have one available (see VIEWING and SEARCHING IN-HOUSE DATABASES, DATABASE SUBSETS and Previous CSD Versions, page 145).
- Select filters (i.e. additional search requirements, e.g. reject disordered structures, reject polymers) (see FILTERS: SECONDARY SEARCH CRITERIA, page 143).
- Set advanced options (currently, the only one available is to standardise hydrogen-atom positions).
- **Cancel**, if **Search** was hit by accident.
- Select **Reset** to return to the status of the *Search Setup* dialogue box when first displayed.
- Hit **Start Search** to run the search.

2.12 Viewing Search Results

Structures found by a search can be viewed on the *View Results* screen.

2.12.1 Monitoring Progress During a Search

- As structures are found, their refcodes are transferred into a list on the right-hand side of *View Results* (see Selecting Individual Structures for Viewing, page 30).
- Structures can be viewed as soon as they are found.
- The progress of the search is indicated by a bar showing the percentage of the database that has been searched. The number of hits found so far is also shown:



2.12.2 Selecting Individual Structures for Viewing

- Any structure found by a search can be selected for viewing – whether or not the search is still running – by clicking on its Refcode (entry ID) (see page 238) in the list on the right-hand side of *View Results*. Alternatively, type the code into the box above the list.
- Only one structure can be viewed at a time.
- The structure currently on display is highlighted in orange:

- The buttons labelled << and >> can be used to step through the structures, backwards or forwards, one by one.
- *Caution* – clicking on the green tick next to an entry in the hit list will suppress that entry (see Suppressing Unwanted Entries, page 40). This means it will not be written out when results are saved in some file formats.

2.12.3 Viewing Information for Individual Structures

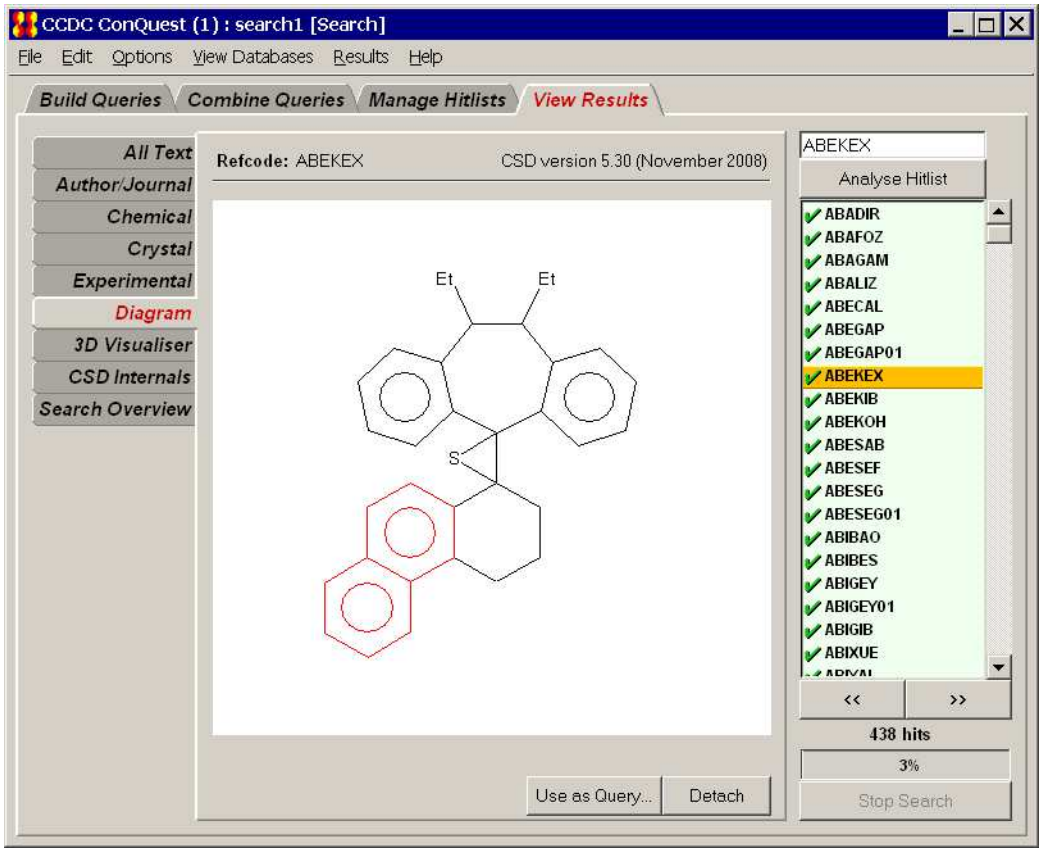
Information for the currently selected entry (i.e. the one highlighted in orange in the *View Results* hit list) is divided into sections. Each section can be displayed by clicking on the appropriate tab at the left-hand side of the *View Results* screen. To detach the selected results pane into its own window click on the **Detach** button. This enables the selected pane and any other panes in *View Results* to be viewed simultaneously. The following results panes are available:

- *All Text*: All CSD information fields that are displayed under the *Author/Journal*, *Chemical*, *Crystal*, and *Experimental* tabs (see below). Note that not all of the possible CSD information fields are present in all entries.

- *Author/Journal:* bibliographic information including deposition number where available (see Deposition Numbers, page 224).
  - A deposition hyperlink may be displayed for recent CCDC numbers (200000 or greater). The hyperlink indicates that original supplementary structure data have been deposited in the CCDC electronic archive. These CIFs may contain data not included in a CSD entry. They have not been subjected to editorial checks. The CIF can be obtained by following the hyperlink to the *CCDC CIF Repository Request Form*: *www.ccdc.cam.ac.uk/products/csd/request/*.
  - If an entry does not contain coordinates and another publication by the same author(s) of the same data set exists with coordinates, a hyperlink is provided in *Notes* to this entry. However if no hyperlink is present from the CCDC deposition number of that entry, coordinates can be obtained by exporting a CIF from ConQuest (see Saving Individual Structures and Lists of Structures, page 41).
  - A **CrossRef** button will be present when it is possible to link through via CrossRef to the original citation.
  - An **IUCr** button will be present when it is possible to link from the CSD entry to IUCr's *Crystallography Journals Online* where the original article will be available.
- *Chemical:* compound name, formula, colour, melting point, etc, plus additional chemical property information where provided by the publishing author.
- *Crystal:* unit cell and space group information. Further information such as the solvent(s) of crystallisation, crystal habit, polymorph information is also provided where supplied by the publishing author.
- *Experimental:* information about experimental conditions, accuracy and precision. Extra details are provided where supplied by the publishing author. CSD editorial notes or comments arising from the data processing are also included e.g. discrepancies between calculated and reported formula weights, discrepancies with density etc.
- *Diagram:* chemical structural diagram. If the search has been for a substructure, then the substructure will be highlighted in the diagram (see Viewing the 2D Chemical Diagram, page 31).
- *3D Visualiser:* a 3D view of the molecule, providing some basic display options (see Using the 3D Visualiser, page 33).
- *CSD Internals*: additional information relating to the entry.
- *Search Overview:* a summary of the query (or queries), filter(s), etc.

2.12.4 Viewing the 2D Chemical Diagram

Each entry in the CSD has a 2D chemical structure diagram associated with it (see Chemical Diagram, page 217) which can be seen in the *Diagram* tab of *View Results* screen. If the search has been run for a substructure, then the substructure will be highlighted:



- If **Detach** is selected the 2D diagram will be detached into its own window enabling the 2D diagram and any other panes in *View Results* to be viewed simultaneously.
- Selection of **Use as Query...,** from either the detached window or from the *Diagram* tab, allows you to use the 2D chemical diagram of the current hit as the starting point for a new search. A *Use-as-Query Options* dialogue box is generated for the query allowing you to:
  - Include/exclude hydrogen atoms.
  - Select only the largest chemical unit.
  - Make each chemical unit into a separate query.
  - Include all chemical units in a single query.
- The 2D image can be copied and pasted into other Windows applications. ***Please note that this is a Windows-only feature***. Ensure the complete diagram is visible on the screen and select **Edit** from the top level menu, then click on **Copy**. Alternatively the keyboard shortcut *Ctrl+C* can be used.

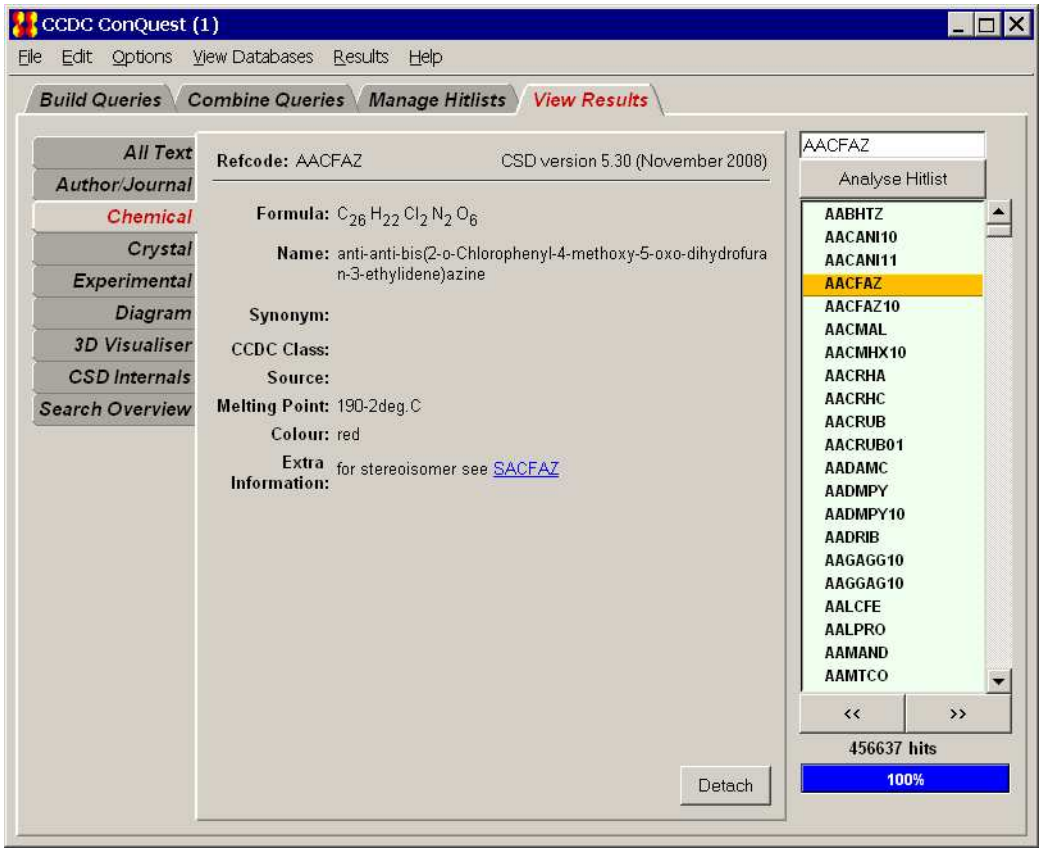
2.12.5 Setting the Default Results Pane

Information for the currently selected entry (i.e. the one highlighted in orange in the *View Results* hit list) is divided into sections. Each section can be displayed by clicking on the appropriate tab at the left-hand side of the *View Results* screen.

By default the 2D diagram is the first results pane displayed when viewing an entry. To change this select **Options** from the top level menu, **Default Results Pane** from the resulting menu, and choose the pane you want displayed by default when first viewing an entry.

2.12.6 Hyperlinking Between Entries

If an entry contains a cross-reference to another CSD entry (identified by a Refcode (entry ID) (see page 238) in, e.g. the *Chemical* and *Experimental* tabs (see Viewing Information for Individual Structures, page 30) then it is possible to *hyperlink* to that entry by clicking on the refcode. For example, entry AACFAZ includes the following information in the *Chemical* pane of the *View Results* screen:



The refcode (entry ID) is shown in blue and underlined. By clicking on this, ConQuest will jump, or hyperlink, to the entry SACFAZ. If the refcode (entry ID) is included within the current hitlist then it will move to this entry; if the refcode is not included in the current hitlist then a new ConQuest browser window will be opened showing the cross-referenced entry together with any other studies of the same chemical compound (see Refcode (entry ID), page 238).

2.13 Using the 3D Visualiser

The *3D visualiser* will allow you to:

- Rotate, translate and scale molecules (see Rotating, Translating and Scaling, page 34).
- Select from a variety of display styles and specify whether hydrogen atoms are to be displayed (see Changing Display Styles and Display of Hydrogen Atoms, page 34).
- Measure molecular geometry: distances, angles and torsions (see Measuring Distances, Angles and Torsions, page 35).
- Label atoms (see Labelling Atoms, page 36).
- Display crystallographic unit-cell contents (see Displaying Crystallographic Unit-Cell Contents, page 36).
- Highlight the hit fragment and show the results of nonbonded contact searches (see Highlighting Hit Fragments and Showing Results of Nonbonded Contact Searches, page 36).
- Display 3D parameters and objects (see Displaying 3D Parameters and Objects, page 37).
- Write out images as *.jpeg* files (see Saving Visualiser Images, page 38).
- Save your current visualiser settings for use in subsequent ConQuest sessions (see Saving Visualiser Settings, page 38).

2.13.1 Visualiser Basics

- Hit *3D Visualiser* in *View Results* to see the currently selected structure in 3D.
- Use the mouse buttons to move the molecule.
- Right-click anywhere in the visualiser area to produce the menu for measuring geometry, changing display styles and generating packing diagrams.

2.13.2 Rotating, Translating and Scaling

- **Movement:** Rotate by moving the cursor around in the 3D window while keeping the left-hand mouse button pressed down. Rotate around the z-axis (the axis perpendicular to the screen) by keeping the left-hand mouse button *and* the Shift key pressed down. Translate by moving the cursor in the 3D window with the centre button depressed (requires 3-button mouse).
- **Scale**: Zoom in or out by moving the cursor up and down in the 3D window while keeping the right-hand mouse button pressed down.
- **Reset***:* Right-click anywhere in the visualiser area and select **Reset View** to set the scale and orientation back to the default.
- If you are using a display style other than wireframe you can increase the speed of rotation, translation or scaling by selecting **Wireframe During Move** from **Settings**; the display style will automatically switch to wireframe when you move the molecule.

2.13.3 Changing Display Styles and Display of Hydrogen Atoms

- A choice of four display styles is offered: wireframe, capped stick, ball-and-stick and space-

filling. Also, the display of hydrogen atoms can be switched on or off.

- To change the display style of the whole molecule or unit cell, right click anywhere in the visualiser box. Select **Display Style** from the resulting pull-down menu and then pick the required style.
- To change the style of just a few atoms, select them by clicking on them with the left-hand mouse button while keeping the Shift key pressed down. Then right-click anywhere to get the pull-down menu, hit **Display Style** and pick the required style.
- To increase the thickness of the lines used for the bonds in wireframe style, right-click anywhere in the visualiser area to get the pull-down menu, select **Settings** and pick **Thicken Wireframe**.
- Right-click anywhere in the visualiser area to get the pull-down menu, select **Settings** and pick **Show Circles for Aromatics** to distinguish aromatic rings.
- Drag the right-hand mouse button while keeping the Shift key depressed to change the *field of view*, i.e. to go from orthographic projection (the default) to perspective projection with increasingly large viewing angles.
- Right-click anywhere in the visualiser area and select **Display Hydrogens** to switch the display of hydrogen atoms on and off.
- Right-click anywhere in the visualiser area and select **Display Disordered Atoms** from the pull-down menu to control the display of atoms that have been identified as being disordered atoms. Specifically, the display of any atoms that have been suppressed by the CSD editors can be controlled using this feature.

2.13.4 Measuring Distances, Angles and Torsions

There are two ways of measuring distances, angles and torsions:

Method 1

- Right-click anywhere in the visualiser area, select either **Measure** or **Picking Mode** from the resulting pull-down menu, then select either **Pick Distances, Pick Angles** or **Pick Torsions**. Depending on which mode has been chosen, you can then click on two, three or four atoms, respectively, to measure a distance, angle or torsion.
- You will remain in the chosen measurement mode, so after measuring the first distance (or angle or torsion), you can continue measuring others.
- To get out of the measurement mode, right-click and select **Measure** followed by **Default Picking Mode**.

Method 2

- If you are in **Default Picking Mode**, select two, three or four atoms by clicking on them with the left-hand mouse button while keeping the Shift key pressed down. Then, right-click and select **Measure** from the pull-down menu. Depending on how many atoms were selected (2,3 or 4) you will be able to click on **Distance, Angle** or **Torsion** to measure the appropriate parameter.
- To clear measurements from the display, right-click and select **Clear Measurements**.

2.13.5 Labelling Atoms

- Right-click anywhere in the visualiser area and select **Label**. You will be given the choice of labelling no atoms (**No Labels**), all atoms (**All Atoms**), all atoms except C, H, N and O (**No C,H,N or O**), all atoms except C and H (**No C or H**), all atoms except H (**No H**), or selected atoms (**Selected**).
- If you choose **Selected**, click on the particular atoms you want labelled.

2.13.6 Displaying Crystallographic Unit-Cell Contents

- Right-click anywhere in the visualiser area, select **Packing** from the resulting pull-down menu and then **1 Unit Cell** from the menu that follows.
- To switch off the display of the unit cell axes right-click anywhere in the visualiser area, select **Packing** from the resulting pull-down menu and disable the **Show Cell Axes** option in the menu that follows. This option to display the unit cell axes is only available when packing is shown.
- To go back to displaying the molecule, right-click in the visualiser area and select **Packing** followed by **Molecule.**

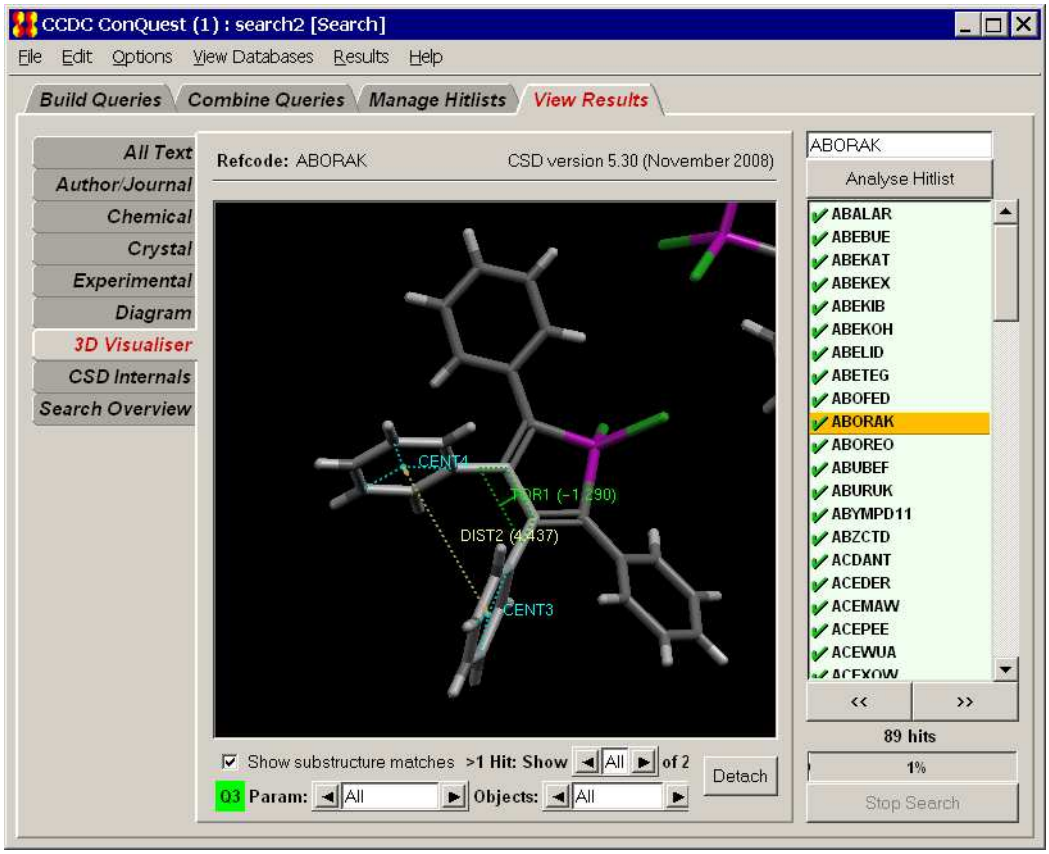
2.13.7 Highlighting Hit Fragments and Showing Results of Nonbonded Contact Searches

- If the check-box labelled **Show substructure matches** is switched on, the visualiser will highlight the hit fragment, i.e. the atoms in the 3D structure that match those of the query substructure. This check-box may be found directly underneath the main visualiser display area.
- If the search was for nonbonded contacts, the contacts will be displayed in the visualiser if (and only if) the **Show substructure matches** check-box is switched on.
- The highlighting of the hit fragment can be made more or less obvious by right-clicking anywhere in the visualiser area, selecting **Highlight Hit** from the resulting pull-down menu and then selecting **Normal Contrast**, **More Contrast** or **Less Contrast**, as desired. Alternatively, it can be switched off altogether by selecting **No Contrast**; this is useful if you want to display nonbonded contacts without highlighting.
- Rather than highlighting the hit fragment, it can be shown in other ways by right-clicking anywhere in the visualiser area, selecting **Highlight Hit** from the resulting pull-down menu and then selecting **Hit as Capped Stick,** which displays the hit fragment in capped-stick style (only available if the current display style is wireframe), or **Single Colour,** which displays the hit fragment coloured by element type and the rest of the structure in one colour.
- If an entry contains more than one hit fragment (i.e. the query substructure can be matched onto the entry in two or more different ways), this will be indicated by the presence of the words *Which match*: beneath the visualiser display area. Clicking on the arrows next to this text shows each substructure match in turn, or all of them at the same time. This option is only available if a 3D or nonbonded contact search has been done, otherwise, only one, arbitrarily chosen, hit fragment is shown, even if the entry contains more than one.

- Hit fragment highlighting is transferred to Mercury when View in Mercury is selected, see (see page 45). For more information about hit highlighting manipulation in Mercury, please refer to the Mercury documentation.

2.13.8 Displaying 3D Parameters and Objects

- If 3D geometric parameters or objects were defined in a substructure search (see Geometric Parameters, page 82), and the check-box labelled **Show substructure matches** is switched on, the visualiser will highlight the hit fragment and display the values of the defined parameters and objects:



- The arrows on either side of the boxes labelled **Param** and **Objects** can be used to control whether or not individual distances, angles, centroids and planes etc. are displayed.
- To change the display style of the parameter or object, right-click anywhere in the visualiser area to get the pull-down menu, select **Settings** and either **Show Parameters as Solid** or **Show Objects as Solid** from the resulting pull-down menu.
- If you are using a display style other than wireframe you have the option of making the hit fragment transparent; select **Settings** and **Make Hit Transparent** from the resulting pull-down menu. This can aid the visualisation of objects which would otherwise be obscured.

2.13.9 Saving Visualiser Images

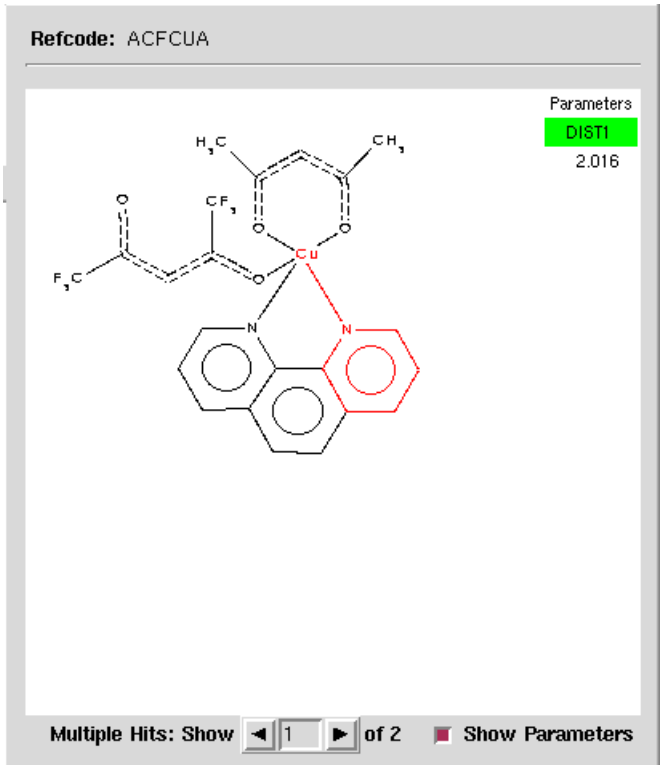
- An image displayed in the 3D visualiser can be saved by right-clicking anywhere in the visualiser area and selecting **Save as JPEG...** from the resulting pull-down menu.
- Alternatively, the 3D image can be copied and pasted into other Windows applications. ***Please note that this is a Windows-only feature***. Ensure the complete diagram is visible on the screen and select **Edit** from the top level menu, then click on **Copy**. The keyboard shortcut *Ctrl+C* may also be used.

2.13.10 Saving Visualiser Settings

- Your current visualiser settings (e.g. of display styles, hit fragment highlighting, etc.) can be saved by right-clicking anywhere in the visualiser area and selecting **Save current setting** from the resulting pull-down menu.
- The effect of this will be that the settings you have saved will be used by default in future ConQuest sessions. You will be able to change them again if you wish.

2.14 Displaying the Values of Defined Geometrical Parameters

- If 3D geometric parameters were defined in a substructure search (see Geometric Parameters, page 82), their values in a hit structure can be seen by moving to the *Diagram* display in *View Results* and turning on the check button labelled **Show Parameters**.
- If the query substructure occurs more than once in a particular entry, this will be indicated by the presence of the words *Multiple Hits: Show* beneath the diagram. Clicking on the arrows next to this text shows each substructure match in turn, along with the values of the associated 3D parameters:


Refcode: ACFCUA
Parameters
DIST1
2.016
Cu
N
N
Multiple Hits: Show
1
of 2
Show Parameters

2.15 Switching between Searches

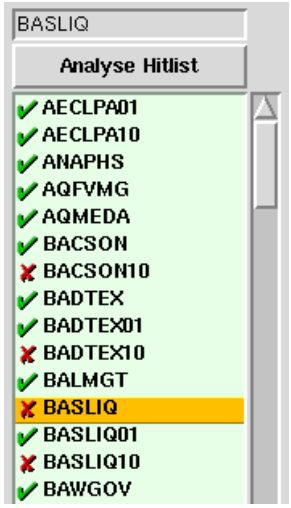
- Any number of searches may be run in a single ConQuest session. New searches can be started while an existing search is still running.
- By default, *View Results* will display the latest search to be started.
- To switch between searches, click on **Results** in the top-level menu and select the required search from **Searches** in the resulting pull-down menu. The currently-selected search (i.e. the one currently displayed in *View Results*) will be greyed out in this menu.
- A star next to a search means that it has not been saved, or has been modified (i.e. entries suppressed or unsuppressed (see Suppressing Unwanted Entries, page 40) since it was last saved.

2.16 Cancelling and Stopping Searches

- A search may be cancelled by hitting **Cancel** in the *Search Setup* dialogue box (see Naming Searches and Setting Filters, page 28) or stopped by hitting the button labelled **Stop Search** on the *View Results* screen.
- Once stopped, a search cannot be resumed from the position at which it was stopped. However, it can be saved and retrieved (see Saving and Reading Searches, page 41) in the same way as searches that have run to completion.

2.17 Suppressing Unwanted Entries

- Clicking on the green tick next to a Refcode (entry ID) (see page 238) in the *View Results* hit list will *suppress* that structure (indicated by a red cross). Four entries are suppressed in the example below:



- Suppressed structures will *not be written out* if results from the search are saved in any format other than the native ConQuest *.cqs* format (see Saving and Reading Searches, page 41). However, they can still be viewed in ConQuest.

- Click on the red cross to reinstate a suppressed structure.
- To suppress all entries, right-click anywhere in the box containing the hit list and pick **Deselect All** from the resulting pull-down menu.
- To unsuppress all entries (i.e. reinstate any that you have suppressed), right-click anywhere in the box containing the hit list and pick **Select All** from the pull-down menu.
- To reverse the suppression status of all entries, right-click anywhere in the box containing the hit list and pick **Invert Selection**.

2.18 Saving and Reading Searches

- To save a search, ensure that it is currently on display in *View Results* (if it is not, display it by using the **Results** top-level menu option, then select **Searches** from the resulting pull-down menu to display the search of your choice (see Switching between Searches, page 40).
- Then pick **File** from the top-level menu and **Save Search** or **Save Search as...** in the resulting pull-down menu. The first of these options will only be available if the search has already been saved (in which case, the only point in saving it again would be if one or more hit entries had been suppressed or unsuppressed since the last save (see Suppressing Unwanted Entries, page 40).
- Search files will be given the extension *.cqs*.
- Search (*.cqs*) files are binary and contain all the information necessary to read the search back into ConQuest and display it as if it had just been run.
- To read a previously saved search, pick **File** from the top-level menu followed by **Open** and then select **Search...** from the resulting pull-down menu.
- Alternatively, you can read in a search, Refcode list or hitlist combination, saved from an earlier ConQuest session, by picking **File** from the top-level menu followed by **Open** and then select **Recent Files** from the resulting pull-down menu.

2.19 Saving Individual Structures and Lists of Structures

Pick **File** from the top-level menu followed by **Export Entries as...** in the resulting pull-down menu to save:

- the individual structure currently being displayed (turn on **Current entry only**), or
- all unsuppressed structures in the hit list (turn on **All selected entries**) (see Suppressing Unwanted Entries, page 40).

Available formats are:

- **CIF** (see CIF Format, page 218): International Union of Crystallography format for crystal structures. The check button **Additional CIF data items** will be switched on by default. Switch on the check button for **Structure data only** if you do *not* wish to export additional text and numeric

fields. When saving all the structures in the hit list, use the buttons to control whether each structure is written to a separate file or all structures are written one after another in a single file. It is also possible to export information on connectivity and defined geometric parameters. To export bond connectivity information switch on the check button **Include Bonds**. This will allow _geom_bond[ ] data items to be included in the CIF.To export data calculated via the use of defined geometric parameters switch on the check button **Include defined parameters**. This will allow the inclusion of _ccdc_geom[ ] data items in the CIF. Distance, angle and torsion data items can all be saved.It is not possible to link parameters derived from centroids, planes or those resulting from combinations back to atom site labels.

- **CIFMIF** (see CIFMIF Format, page 219): This format combines CIF crystallographic data names with chemical data names from the Molecular Information File (CIFMIF has additional non-standard fields containing CSD-specific information, e.g. information about whether the structure has unresolved errors). It is possible to export data calculated through the use of defined geometric parameters by switching on the **Output defined parameters** check button. For more details refer to CIF export (above).
- **COORD** (see COORD Format, page 221): Use the buttons to control whether fractional or orthogonal atomic coordinates are written out, and whether the complete structure is required or just those atoms in the structure that match the search substructure.
- **TSV**: Tab separated values. This format can be used to import references into the EndNote software tool for publishing and managing bibliographic data (see *http://www.endnote.com/*).
- **FDAT** (see FDAT Format, page 227): CCDC format for crystal structures; necessary for transferring structures into Pluto (see page 236).
- **MOL2** (see Mol2 Format, page 233): Tripos Inc. format for 3D molecules and crystal structures. Switch on the check-box labelled **Normalise H coordinates** if you want C-H, N-H and O-H hydrogen atom positions standardised (the H atom will be moved along the experimentally determined bond vector to make the bond length equal to the average neutron diffraction value (see Hydrogen Atom Location in Crystal Structure Analyses, page 228). The **Use CCDC bond types** check-box can be used to specify whether or not CCDC bond type assigmants are written to output files (see Bond Type Conventions, page 216). Switch on the check box labelled Export largest molecule only if you wish to only write out the largest distinct chemical unit (i.e. molecule) in the structure (see Molecule, page 233). When saving all the structures in the hit list, use the buttons to control whether each structure is written to a separate file or all structures are written one after another in a single file.
- **PDB** (see pdb Format, page 235): Protein Data Bank format for 3D molecules. When saving all the structures in the hit list, use the buttons to control whether each structure is written to a separate file or all structures are written one after another in a single file.
- **Refcode** (see Refcode (entry ID), page 238): simple ASCII refcode list (default extension *.gcd*).
- **SD** (see SD Format, page 240): MDL Information Systems, Inc. format for 3D molecules. The **Use 'Any' bond type** check-box should be switched on if you want CSD bond types that are

unrecognised in the MDL conventions to be set to the MDL *Any* bond type; otherwise, they will be written out as single bonds. (NB. Some third-party programs cannot read *.sd* files containing molecules with bonds of type *Any*.) The **Exclude any disordered atoms** check-box can be used to specify whether or not disordered atoms should be written out. Switch on the check-box labelled **Normalise H coordinates** if you want C-H, N-H and O-H hydrogen atom positions standardised (the H atom will be moved along the experimentally determined bond vector to make the bond length equal to the average neutron diffraction value (see Hydrogen Atom Location in Crystal Structure Analyses, page 228). Switch on the check box labelled Export largest molecule only if you wish to only write out the largest distinct chemical unit (i.e. molecule) in the structure (see Molecule, page 233).

- **SHELX:** simulated SHEL-X *.res* file.
- **TAB:** Tab separated list of text items. Click on the check buttons next to individual text items (Bibliography, Chemical, etc.) to indicate which are to be included.
- **TXT***:* Plain text format. Click on the check buttons next to individual text items (Bibliography, Chemical, etc.) to indicate which are to be included.

2.20 Transferring Data to Other Programs for Analysis

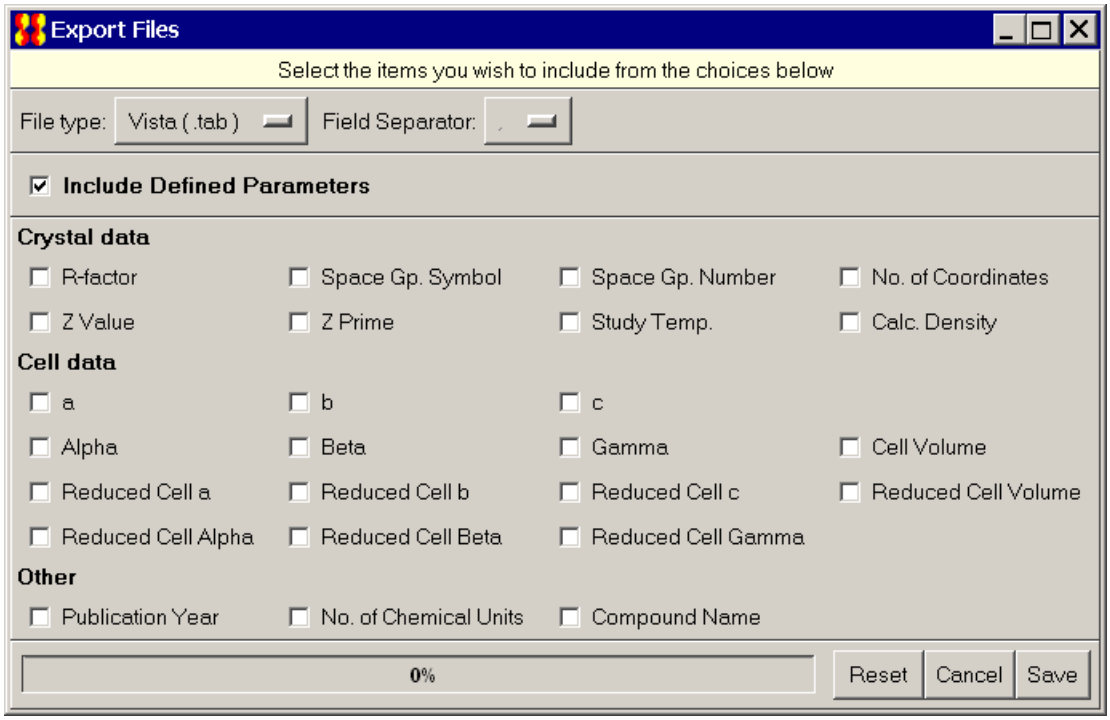
2.20.1 Transferring Results Directly to Vista or Excel

- Depending on whether ConQuest is being run under Unix or Windows, data fields and 3D parameters for unsuppressed hit structures (see Suppressing Unwanted Entries, page 40) can be transferred to Vista and/or Excel for statistical analysis. Select **File** from the top-level menu and then **View in Vista...** or **View in Excel...** from the resulting pull-down menu. Alternatively, right-click in the box containing the hitlist and pick **View in Vista...** or **View in Excel...** from the resulting pull-down menu.
- This generates a dialogue box from which data fields and parameters can be selected.
- Hit **View in Vista (**or **View in Excel)** to transfer data to Vista (or Excel).

2.20.2 Saving Data for Later Analysis in Vista, Excel, etc.

Data fields and 3D parameters for unsuppressed hit structures (see Suppressing Unwanted Entries, page 40) can be saved to a file for later analysis:

- Select **File** from the top-level menu and then **Export Parameters and Data...** from the resulting pull-down menu; this generates the *Export Files* dialogue box:



- Select the *File type* you wish to save:
  - **Vista** (*.tab*) for viewing data in Vista.
  - **Spreadsheet** (*.spr*) to save a file where data items for each entry are on a single line separated by a character selected from the *Field Separator* menu. This allows data to be saved for input to spreadsheet packages such as Excel.
  - If 3D parameters were defined in a substructure search the option of saving a legacy Quest3D Summary file is also available (**QUEST Summary** (*.sum*)).
- Select data fields and parameters of interest and hit **Save.**

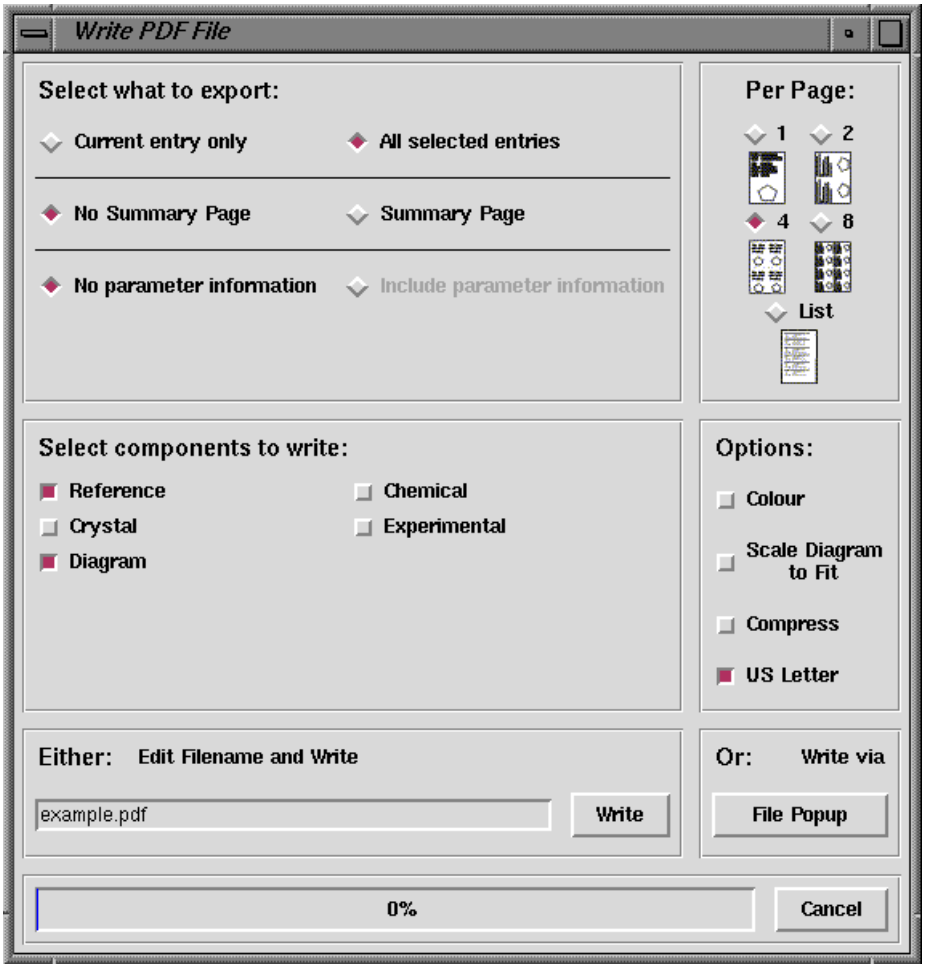
2.21 Viewing Results in Mercury

Mercury (see page 232) is an advanced crystal structure visualiser distributed as part of the CSD system for viewing, either your own crystal structures, or those retrieved from a ConQuest search. To view the hits from a ConQuest search in Mercury hit **Analyse Hitlist** and select **View in Mercury** from the pull-down menu. Alternatively, select **File** from the top-level menu and **View in Mercury** from the resulting pull-down menu.

2.22 Printing Search Results

Search results (diagrams, literature references, etc.) can be written out as *.pdf* files, which may then be viewed and printed with Adobe Acrobat Reader (included on your ConQuest CD-ROM and also freely available from *http://www.adobe.com*). Acroread on UNIX has a *-toPostscript* option which converts PDF to Postscript and could be used to save postscript.

- In *View Results*, select **File** from the top-level menu and **Write PDF file to view/print...** from the resulting pull-down menu.
- Use the buttons in the *Write PDF File* dialogue box to control the contents of the PDF file, e.g. number of entries per page, which items (diagram, journal reference, etc.) are to be included, whether information is to be written for the structure currently on display or for all unsuppressed structures in the hit list (see Suppressing Unwanted Entries, page 40).
- For example, the setting below will result in the diagrams and literature references being written out for all unsuppressed structures in the hit list, 4 per page, in US Letter format, and with no summary page. The file will be called *example.pdf*.


Write PDF File
Select what to export:
Current entry only
All selected entries
No Summary Page
Summary Page
No parameter information
Include parameter information
Per Page:
1
2
4
8
List
Select components to write:
Reference
Chemical
Crystal
Experimental
Diagram
Options:
Colour
Scale Diagram to Fit
Compress
US Letter
Either: Edit Filename and Write
example.pdf
Write
Or: Write via
File Popup
0%
Cancel

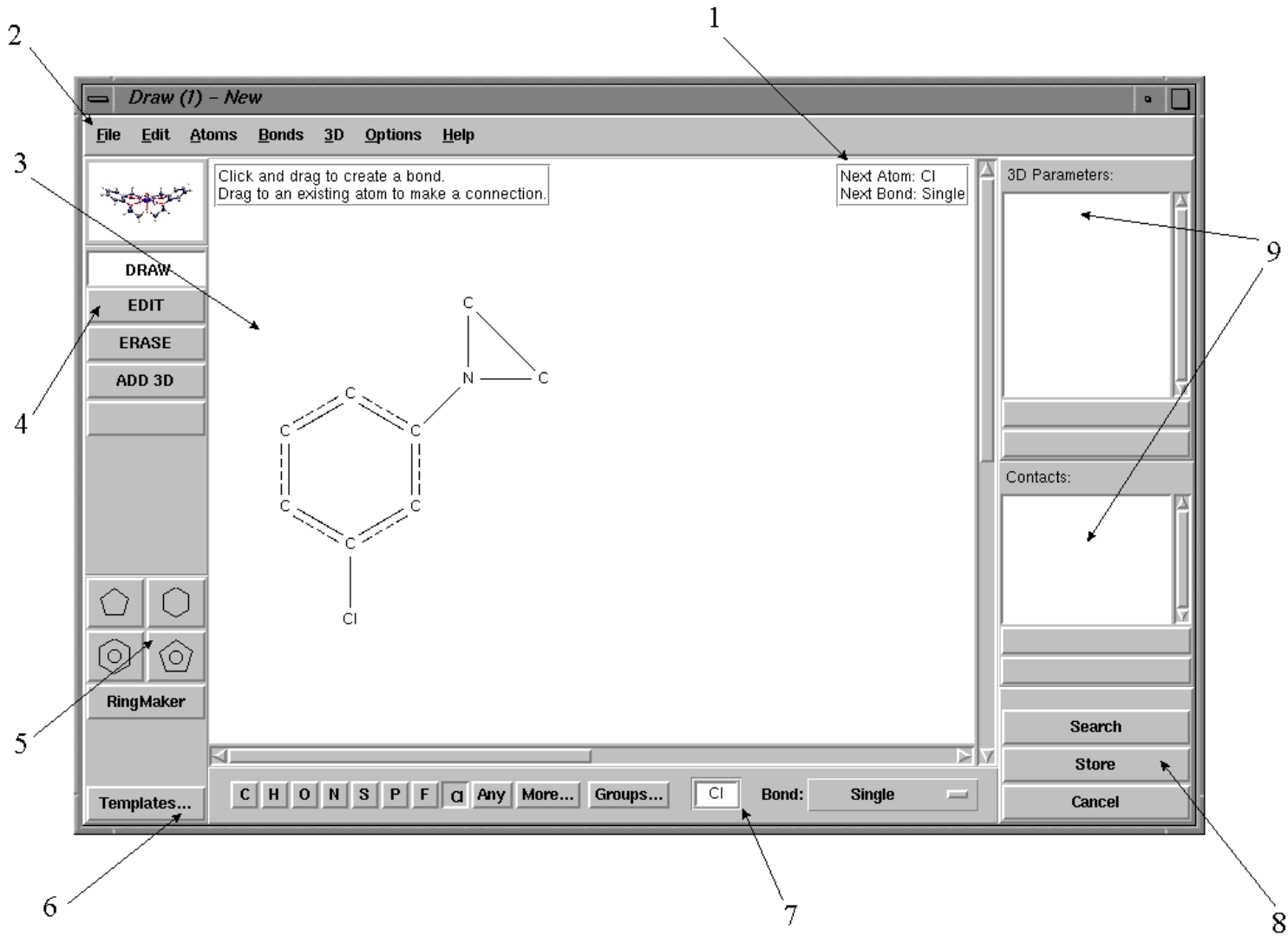
3 SUBSTRUCTURE SEARCHES

3.1 Typical Uses of Substructure Searching

- Substructure searching, e.g. find all 8-membered carbocycles.
- Finding particular molecules, e.g. find caffeine.
- As the essential preliminary to geometrical studies and 3D searching.

3.2 Layout of the Draw Window

Open the substructure drawing window by hitting **Draw** in *Build Queries*.



1. Help messages.
2. Top-level menu (different from the menu in the main ConQuest interface).
3. Drawing area (see Fundamentals of Drawing, page 49).
4. Mode buttons (see Modes in the Draw Window, page 49) – *responses to mouse clicks in the drawing area will depend on which mode is active*.

5. Area for selecting basic ring templates (basic carbon rings to aid drawing (see Adding a Ring to a Blank Drawing Area, page 52).
6. Area for selecting other templates (molecular building blocks to aid drawing) (see Using Chemical Groups and Substructure Templates, page 68).
7. Area for changing the current element type (see Changing the Current Element Type, page 56) and bond type (see Changing the Current Bond Type, page 61).
8. Buttons for starting a search, storing a substructure query, or cancelling everything that has been done.
9. Area for listing, displaying and editing 3D (see Geometric Parameters, page 82) and nonbonded contact parameters (see Nonbonded Contacts between Atoms, page 96).

3.3 Modes in the Draw Window

The five buttons down the left-hand side of the *Draw* window are mode buttons which affect what happens when the mouse is used in the drawing area.

- **DRAW**: Click on this button when you want to draw a substructure. The **DRAW** mode button is active if coloured white; activate, if necessary, by clicking on it.
- **EDIT:** Click on this button when you want to perform editing tasks such as moving, rotating or resizing substructures, or selecting atoms or bonds.
- **ERASE**: Click on this button when you want to delete atoms or bonds.
- **ADD 3D**: Click on this button to specify geometric objects, parameters and constraints.
- **CONTACT**: Click on this button to set up a nonbonded contact search (you must already have drawn two unconnected substructures).

3.4 Fundamentals of Drawing

All drawing takes place in the central white area of the *Draw* window.

3.4.1 Drawing a Bond

- Ensure you are in **DRAW** mode.
- Move the cursor into the white area of the *Draw* window.
- Press down the left-hand mouse button, move the cursor *while keeping the mouse button depressed*, and then release the button.
- This draws a bond, using the current element type (see Changing the Current Element Type, page 56) and bond type (see Changing the Current Bond Type, page 61).

3.4.2 Drawing an Isolated Atom

- Ensure you are in **DRAW** mode.
- Move the cursor into the white area of the *Draw* window.
- Click the left-hand mouse button, and release it again without moving the mouse.

3.4.3 Drawing a Bond from an Existing Atom

- Ensure you are in **DRAW** mode.
- Move the cursor onto the atom (the atom will go red).
- Press down the left-hand mouse button.
- Move the cursor while keeping the mouse button depressed, then release the button.

3.4.4 Drawing a Bond to an Existing Atom

- Ensure you are in **DRAW** mode.
- Move the cursor into the white area of the *Draw* window.
- Press down the left-hand mouse button.
- Move the cursor onto the desired atom (the atom will go red) while keeping the mouse button depressed, then release the button.

3.4.5 Drawing a Bond between Two Existing Atoms

- Ensure you are in **DRAW** mode.
- Move the cursor onto the first atom (the atom will go red).
- Press down the left-hand mouse button.
- Move the cursor onto the second atom (the atom will go red) while keeping the button depressed, then release the button.

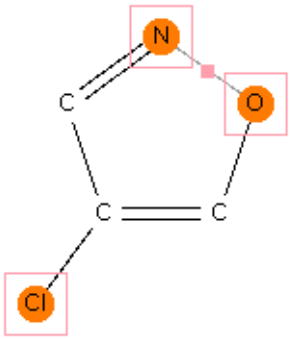
3.4.6 Undoing Mistakes when Drawing Substructures

- Pick **Edit** in the top-level menu and **Undo** in the resulting pull-down menu to undo the last action performed.
- Alternatively, move the cursor to a blank point in the white area, click the right-hand mouse button, and select **Undo** from the pull-down menu.
- If necessary, **Edit**... **Undo** may be used several times in a row to undo a sequence of actions, one by one.

3.4.7 Selecting Atoms and Bonds

- Selection of atoms or bonds is useful for assigning properties such as charge, element type, bond type; for moving substructures around the drawing area; for cutting and pasting.

- A selected atom is coloured orange and enclosed in a box. If the two atoms at either end of a bond are selected then the bond itself is selected too. In the example below, the N, O and Cl atoms and the N-O bond are selected.



Atoms and bonds may be selected in several ways:

- In **EDIT** mode, an individual atom or bond can be selected by clicking on it with the left-hand mouse button.
- In **EDIT** mode, a series of atoms or bonds can be selected by clicking on each in turn while keeping the Shift key pressed down.
- In **EDIT** mode, a group of atoms and bonds can be selected by clicking with the left-hand mouse button on a blank point in the white area and moving the cursor while keeping the mouse button pressed down. Everything enclosed in the resulting rectangular box gets selected when the mouse button is released.
- In any mode, everything can be selected by hitting **Edit** in the top-level menu and **Select All** in the resulting pull-down menu.
- In any mode, the current selection can be reversed by hitting **Edit** in the top-level menu and **Invert Selection** in the resulting pull-down menu. Everything that was selected becomes unselected, and vice versa.

3.4.8 Deleting Atoms and Bonds

There are several methods, including:

- In **DRAW** mode, click with the right-hand mouse button on the atom or bond to be deleted and pick **Delete Atom** or **Delete Bond** from the resulting pull-down menu.
- Change to **ERASE** mode (click on the **ERASE** button) and click with the left-hand mouse button on the atom or bond to be deleted.
- Change to **EDIT** mode (click on the **EDIT** button). Select the atoms or bonds to be deleted (see Selecting Atoms and Bonds, page 50). Then either:

- Pick **Edit** from the top-level menu and then **Cut** or **Delete Selected** from the resulting pull-down menu (**Cut** will delete both atoms and bonds, **Delete Selected** will give a choice).
- Click with the right-hand mouse button on a blank point in the white area and pick **Delete Selected** or **Cut** from the resulting pull-down menu.

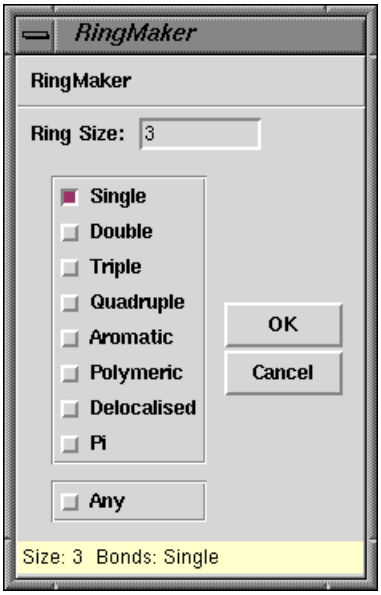
- To delete all atoms and bonds, in **DRAW**, **EDIT** or **ERASE** modes, move the cursor onto a blank point in the white area, click on the right-hand mouse button, and pick **Clear All** from the pull-down menu.
- Use **Edit**... **Undo** (see Undoing Mistakes when Drawing Substructures, page 50) to delete an atom or bond just drawn.

3.5 Drawing and Fusing Rings

3.5.1 Adding a Ring to a Blank Drawing Area

- Rings may be drawn manually but the easiest way is to use the pre-drawn rings in the bottom left-hand corner of the *Draw* window:

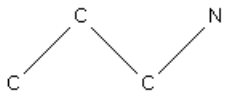
- If the desired ring is one of the four on display (see above), select it by clicking on its icon, move the cursor into the white area, then click with the left-hand mouse button.
- If you click again, you will create a second copy of the ring. Use **Edit... Undo** if this is not what was wanted. Click on **DRAW** to resume normal drawing.
- If a different ring size is required, hit **RingMaker**, type the desired ring size into the box, select the required bond type, hit **OK**, then click in the white area. In the example below, a saturated cyclopropane ring has been defined:



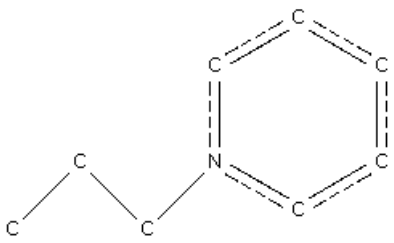
- Only carbon rings can be specified, but it is easy to change individual element types later (see Changing the Current Element Type, page 56).
- Some complex ring systems (e.g. crown ethers) are available by clicking on **Templates...** in the bottom left-hand corner.

3.5.2 Adding a Ring to an Atom in an Existing Substructure

- Select the ring (see Adding a Ring to a Blank Drawing Area, page 52), then click on the desired atom in the existing substructure with the left-hand mouse button.
- For example, selecting a 6-membered aromatic ring and clicking on the N atom in:

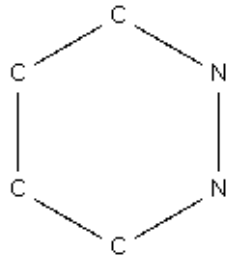


will create:

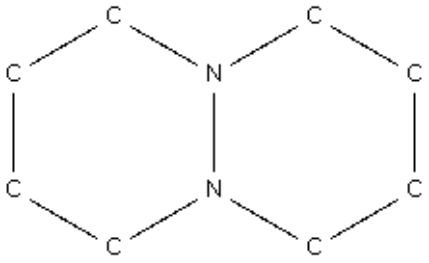


3.5.3 Fusing a New Ring to an Existing Ring

- Select the new ring (see Adding a Ring to a Blank Drawing Area, page 52), then click on the desired fusion bond in the existing ring.
- For example, selecting a 6-membered saturated ring and clicking on the N-N bond in:

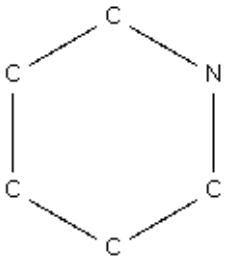


will create:

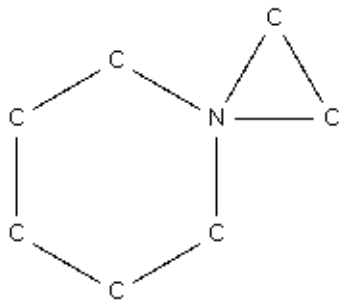


3.5.4 Creating a Spiro-Fusion

- Select the required ring (see Adding a Ring to a Blank Drawing Area, page 52), then click on the desired spiro atom in an existing ring.
- For example, selecting a 3-membered saturated ring and clicking on the N atom in:



will create:



3.5.5 Fusing Rings by Moving One Ring onto Another

It is possible (but sometimes tedious and not recommended) to fuse two separate rings in the white area by selecting all the atoms in one ring (see Selecting Atoms and Bonds, page 50) and moving it towards the other (see Moving Atoms, page 73):

- Spiro fusion is achieved by overlapping one atom in the moveable ring with one atom in the stationary ring (indicated by the overlapped atoms going red). Fusion will occur when the mouse button is released.
- Bond fusion is achieved by overlapping two bonded atoms in the moveable ring with two bonded atoms in the stationary ring. It may be necessary to overlap one of the pairs and then rotate the moveable ring (by holding down the Control key) until the second pair overlap (see Rotating Substructures, page 74).

3.6 Atom Properties

Atoms may have the following properties:

- Element type (this may be a variable type, e.g. N or O) (see Changing the Current Element Type, page 56).
- Attached hydrogens; these may be generated automatically (see Automatic Addition of Hydrogen Atoms, page 58) or added manually (see Adding Hydrogen Atoms Manually, page 59).
- Formal charge (see Setting Atomic Charges, page 60).
- Coordination number (see Setting Atom Coordination Numbers, page 60).
- Cyclicity status (see Defining Cyclic or Acyclic Atoms, page 61).

3.6.1 Changing the Current Element Type

- The current element type determines the type of any new atom created when drawing. It is displayed in the white box at the bottom of the *Draw* window. In the example below, the current element type is chlorine:

- The current element type may be changed by hitting any of the element symbols at the bottom of the *Draw* window.
- Alternatively, it may be typed into the white box at the bottom of the *Draw* window.
- **Any** means any atom (i.e. an atom of any element type), and is denoted by the symbol **X** in the drawn substructure.
- **More...** displays a pull-down menu. Selecting **Other Elements...** from this menu allows selection of any element in the Periodic Table. Use **Single Pick** mode unless a variable element type is needed (see Setting Variable Element Types, page 56).

3.6.2 Setting Variable Element Types

- Atoms in a substructure may be variable, e.g. F *or* Cl *or* Br *or* I.
- **Any** means any atom (i.e. an atom of any element type), and is denoted by the symbol **X** in the drawn substructure.
- There are a number of pre-defined element groups (e.g. *4B* means any of *Ti, Zr, Hf*). The current element type can be changed to one of the groups by typing the element-group symbol into the white box at the bottom of the *Draw* window.
- The current element type (see Changing the Current Element Type, page 56) may be made variable by hitting the **More...** button (at the bottom of the *Draw* window). This displays a menu from which common variable element types (e.g. **Not Hydrogen**, **Any Metal**, etc.) can be selected. Alternatively, selecting **Other Elements...** from this menu opens up the Periodic Table.

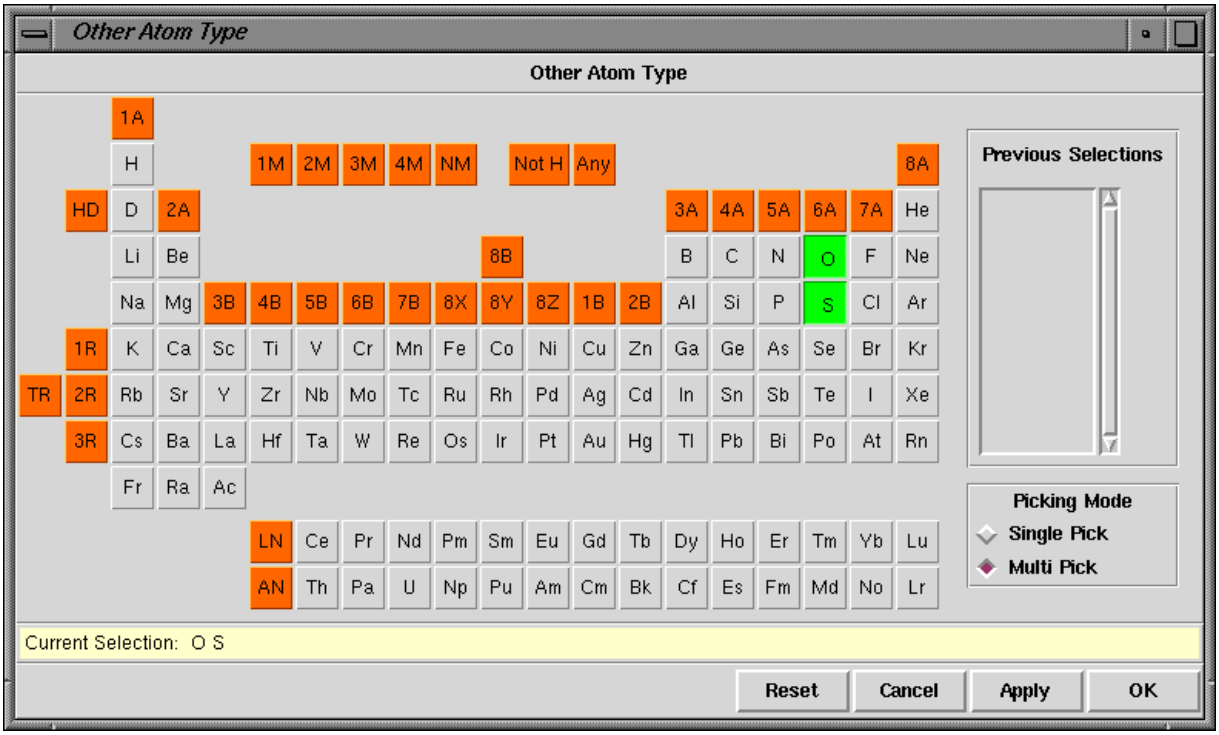
There are then two ways to create a variable element type:

Method 1

- Select **Single Pick** and click on one of the orange pre-defined element groups (see Element Group Symbols, page 225), e.g. *4B* means any of *Ti, Zr, Hf*. Then hit **Apply** or **OK**.

Method 2

- Select **Multi Pick**, click on the required elements, e.g. O and S:



then hit **OK** (closes the window) or **Apply** (leaves the window open). The resulting variable element type is called *QA*, if it is the first variable type created, *QB*, if it is the second, etc.

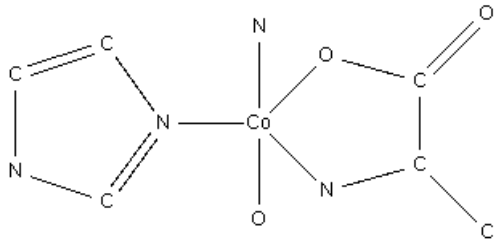
3.6.3 Changing the Element Types of Existing Atoms

This can be done in several ways, including:

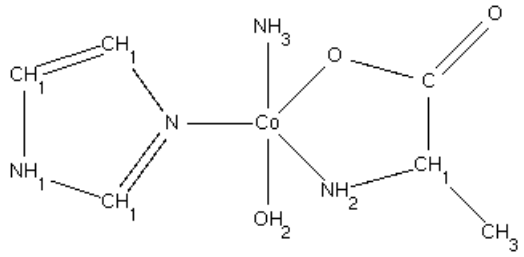
- In **DRAW** or **EDIT** modes, click on the atom with the right-hand mouse button and select **Element** from the resulting pull-down menu. Then select the required element type. **More...** allows selection of some common variable element types or brings up the Periodic Table which, amongst other things, allows creation of customised variable element types (see Setting Variable Element Types, page 56).
- In **DRAW** mode, change the current element type (see Changing the Current Element Type, page 56) and then click on the atom with the left-hand mouse button.
- In **DRAW** or **EDIT** mode, click on **Atoms** in the top-level menu, select **Element** from the resulting pull-down menu and select the required element (**More...** allows selection of some common variable element types or brings up the Periodic Table (see Setting Variable Element Types, page 56). The *Select Atoms* pop-up appears, click on the atom or atoms to be changed with the left-hand mouse button and hit **Done**.
- In **EDIT** mode, select the atom(s) to be changed (see Selecting Atoms and Bonds, page 50). Pick **Atoms** from the top-level menu, **Element** from the resulting pull-down menu, and select the required element. **More...** allows selection of some common variable element types or brings up the Periodic Table, which, amongst other things, allows creation of customised variable element types (see Setting Variable Element Types, page 56).

3.6.4 Automatic Addition of Hydrogen Atoms

Hydrogen atoms may be added automatically to substructures. The number of hydrogens added to each atom is sufficient to satisfy the atom's unfilled valencies. For example, auto-addition of hydrogens to the following substructure:



will create the following:



- To add hydrogens to all atoms, go to **DRAW** or **EDIT** mode, click on **Atoms** in the top-level menu, hit **Hydrogens** in the resulting pull-down menu, then hit **Generate** followed by **All Atoms.**
- To add hydrogens to one or more specified atoms, go to **EDIT** mode, select the required atoms (see Selecting Atoms and Bonds, page 50), click on **Atoms** in the top-level menu, hit **Hydrogens** in the resulting pull-down menu, then hit **Generate** followed by **Selected Atoms**.
- Alternatively, go to **DRAW** or **EDIT** mode, click on **Atoms** in the top-level menu, hit **Hydrogens** in the resulting pull-down menu, then hit **Generate** followed by **Selected Atoms**. Click on the atoms to which you want to add hydrogens and hit **Done**.
- Alternatively, go to **DRAW** or **EDIT** mode, right-click on an atom, select **Hydrogens** from the resulting pull-down menu and **Generate** from the next menu.
- To remove hydrogens, go to **DRAW** or **EDIT** mode, click on **Atoms** in the top-level menu, hit **Hydrogens** in the resulting pull-down menu, then hit **Clear** followed by **All Atoms** or **Selected Atoms**. You can remove hydrogens from a particular atom by right-clicking on the atom and selecting **Hydrogens** and **Clear** from the resulting pull-down menus.

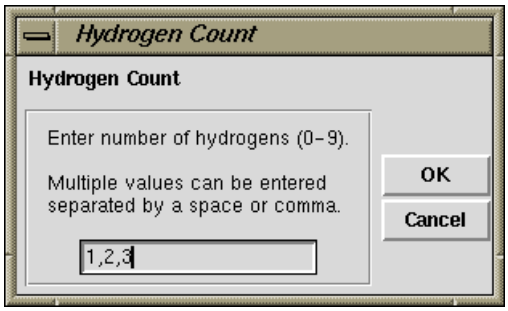
- Occasionally, the program may assign the wrong number of hydrogens to an atom. This is especially likely for hydrogen-bridged metals, the oxygen atoms of metal-coordinated alcohols, and boron atoms in boron cages. It is advisable to check the number of H-atoms added to these types of structures. You can alter or explicitly define the number of hydrogens on an atom by adding them manually (see Adding Hydrogen Atoms Manually, page 59).

3.6.5 Adding Hydrogen Atoms Manually

Hydrogen atoms may be drawn in the same way as any other type of atom or they may be defined implicitly. These two different ways may occasionally give different search results (see Hydrogen Atoms in 2D/3D Substructure Searching, page 227).

Hydrogens may be added implicitly in several ways, including:

- In **DRAW** or **EDIT** mode, click on an atom with the right-hand mouse button, pick **Hydrogens** from the resulting pull-down menu, then select the number of hydrogens required from the second pull-down menu. **Unspecified** in this menu means that any number of hydrogens is allowed. Clicking on **Other...** opens up a dialogue box which allows the specification of numbers greater than 4, or variable numbers of hydrogens. The example below specifies that the atom must be bonded to 1, 2 or 3 hydrogens.

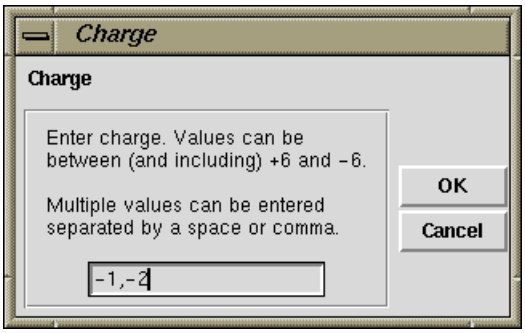


- In **DRAW** or **EDIT** mode, click on **Atoms** in the top-level menu, hit **Hydrogens** in the resulting pull-down menu, then select the required number of hydrogens as above. The *Select Atoms* pop-up appears, click on the atom(s) to which hydrogens are to be added and hit **Done**.
- In **EDIT** mode, select (see Selecting Atoms and Bonds, page 50) the atom(s) to which hydrogens are to be added. Then hit **Atoms** in the top-level menu, **Hydrogens** in the resulting pull-down menu, and pick the required number of hydrogens as above.
- Hydrogen atoms may also be added automatically (see Automatic Addition of Hydrogen Atoms, page 58).

3.6.6 Setting Atomic Charges

The formal (integer) Atomic Charge (see page 212) of an atom may be specified in several ways, including:

- In **DRAW** or **EDIT** mode, click on the atom with the right-hand mouse button, pick **Charge** from the resulting pull-down menu, then select the charge required from the second pull-down menu. **Unspecified** in this menu means that any charge is allowed. **Any+/-** means any non-zero charge. Clicking on **Other...** opens up a dialogue box which allows the specification of variable charges. The example below specifies that the atom must have a charge of *-1* or *-2*.



- In **DRAW** or **EDIT** mode, click on **Atoms** in the top-level menu, select **Charge** from the resulting pull-down menu, then pick the required charge as above. The *Select Atoms* pop-up appears, click on the atom(s) to which the charge is to be assigned and hit **Done**.
- In **EDIT** mode, select (see Selecting Atoms and Bonds, page 50) the atom(s) to which a charge is to be assigned. Then hit **Atoms** in the top-level menu, **Charge** in the resulting pull-down menu, and pick the required charge as above.

3.6.7 Setting Atom Coordination Numbers

It is possible to specify the coordination number of an atom (i.e. the total number of atoms to which it is bonded). This is necessary, e.g., if the atom is bonded to only four atoms in the substructure that has been drawn but is required to be 6-coordinate overall.

Total coordination numbers can be specified in several ways, including:

- In **DRAW** or **EDIT** mode, click on an atom with the right-hand mouse button, pick **Number of Bonded Atoms** from the resulting pull-down menu, then select the coordination number required from the second pull-down menu. **Unspecified** in this menu means that any coordination number is allowed. Clicking on **Other...** opens up a dialogue box which allows the specification of values higher than 8.

- In **DRAW** or **EDIT** mode, click on **Atoms** in the top-level menu, select **Number of Bonded Atoms** from the resulting pull-down menu, then select the required coordination number. Click on the atom(s) to which the coordination number is to be assigned and hit **Done**.
- In **EDIT** mode, select (see Selecting Atoms and Bonds, page 50) the atom(s) whose coordination number is to be defined. Then hit **Atoms** in the top-level menu, **Number of Bonded Atoms** in the resulting pull-down menu, and select the required coordination number.

3.6.8 Defining Cyclic or Acyclic Atoms

It is possible to specify that a particular atom must be cyclic (i.e. part of a ring) or, conversely, that it must be acyclic (i.e. not part of a ring). This can be done in several ways, including:

- In **DRAW** or **EDIT** mode, click on an atom with the right-hand mouse button, pick **Cyclicity** from the resulting pull-down menu, then select the required option from the second pull-down menu. **Unspecified** in this menu means the atom may be either cyclic or acyclic. If the atom is already part of a ring, the **Acyclic** option will be greyed out.
- In **DRAW** or **EDIT** mode, click on **Atoms** in the top-level menu, select **Cyclicity** from the resulting pull-down menu, then select **Cyclic, Acyclic** or **Unspecified**. Click on the atom(s) to which the option is to be applied and hit **Done**.
- In **EDIT** mode, select (see Selecting Atoms and Bonds, page 50) the required atom(s). Then pick **Atoms** from the top-level menu, **Cyclicity** from the resulting pull-down menu, and select **Cyclic, Acyclic** or **Unspecified** as required.

3.7 Bond Properties

3.7.1 Changing the Current Bond Type

- The current bond type determines the type of any new bond created when drawing (see Bond Type Conventions, page 216). The current setting is shown on the button next to the word *Bond* at the bottom of the *Draw* window.

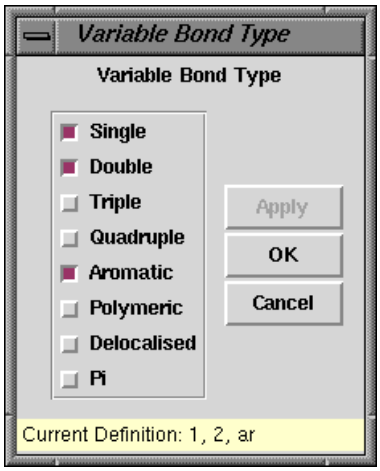
- The current bond type may be changed by clicking on this button and selecting from the resulting pull-down menu.

3.7.2 Setting Variable Bond Types

- Bonds in a substructure may be variable, e.g. double or aromatic (see Variable Bond Type, page 246).
- The current bond type (see Changing the Current Bond Type, page 61) can be made variable by

clicking on the button next to the word *Bond* at the bottom of the *Draw* window. Select **Variable...** from the pull-down menu, click on the required bond types in the resulting pop-up window, then hit **OK** (closes window) or **Apply** (leaves window open).

- The example below shows the setting required to create a variable bond type of single, double or aromatic:



3.7.3 Changing the Types of Existing Bonds

This can be done in several ways, including:

- In **DRAW** or **EDIT** modes, click on the centre of the bond with the right-hand mouse button and select **Type** from the resulting pull-down menu. Then pick the required bond type (this may be **Variable...** (see Variable Bond Type, page 246).
- In **DRAW** mode, change the current bond type and then click on the bond with the left-hand mouse button.
- In **DRAW** or **EDIT** mode, click on **Bonds** in the top-level menu, select **Type** from the resulting pull-down menu, select the required bond type (this may be **Variable...** (see Variable Bond Type, page 246). The *Select Bonds* pop-up appears, click on the bond(s) to be changed with the left-hand mouse button and hit **Done**.
- In **EDIT** mode, select the bond(s), pick **Bonds** from the top-level menu, **Type** from the resulting pull-down menu, then pick the required bond type (this may be **Variable...** (see Variable Bond Type, page 246).

3.7.4 Defining Cyclic or Acyclic Bonds

It is possible to specify that a particular bond must be cyclic (i.e. part of a ring) or acyclic (i.e. not part of a ring). This can be done in several ways, including:

- In **DRAW** or **EDIT** mode, click on the centre of the bond with the right-hand mouse button, pick **Cyclicity** from the resulting pull-down menu, then select the required option from the second

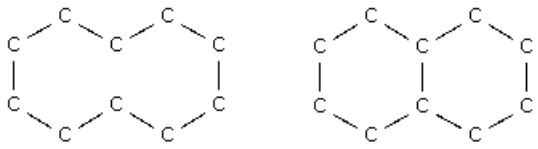
pull-down menu. **Unspecified** in this menu means the bond may be either cyclic or acyclic. If the bond is already part of a ring, the **Acyclic** option will be greyed out.

- In **DRAW** or **EDIT** mode, click on **Bonds** in the top-level menu, select **Cyclicity** from the resulting pull-down menu, then select **Cyclic, Acyclic** or **Unspecified**. The *Select Bonds* pop-up appears, click on the bond(s) to be changed with the left-hand mouse button and hit **Done**.
- In **EDIT** mode, select the required bond(s) (see Selecting Atoms and Bonds, page 50). Pick **Bonds** from the top-level menu, **Cyclicity** from the resulting pull-down menu, then select **Cyclic, Acyclic** or **Unspecified** as required.

3.8 Avoiding Unwanted Bonds and Rings

When drawing substructures, it is possible to:

- Avoid unwanted bonds between two or more atoms in the substructure (see Avoiding Unwanted Bonds between Atoms in a Substructure, page 64), e.g. to specify that a pair of atoms must *not* be bonded to each other. The usual reason for doing this is to avoid links across a ring, e.g. to avoid finding decalin when searching for cyclodecane:

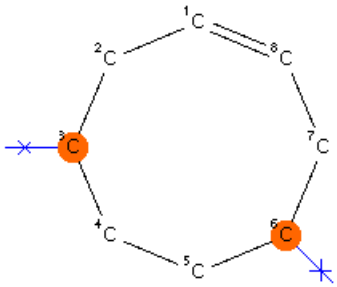


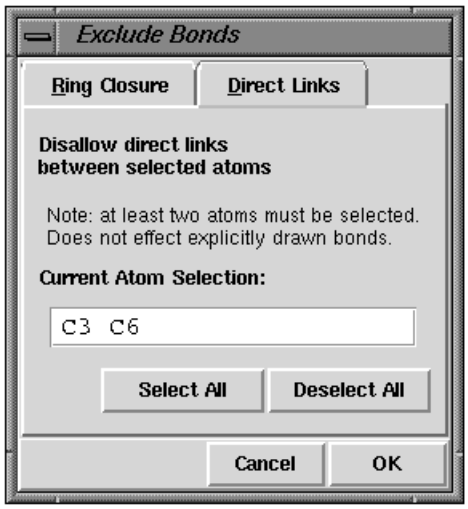
cyclodecane decalin

- Avoid unwanted cyclic bonds to atoms outside the substructure (see Avoiding Unwanted Cyclic Bonds to Atoms outside the Substructure, page 65), e.g. to ensure that there are no rings fused or spiro-linked to the substructure you have drawn.
- Ensure that ring systems in the substructure are *isolated* (see Ensuring that a Ring System has no Additional Bridging or Fusion, page 66), i.e. have no fusion or bridging other than what you have specified.

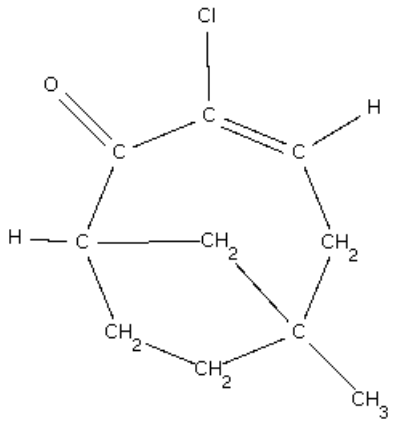
3.8.1 Avoiding Unwanted Bonds between Atoms in a Substructure

- Select **Bonds** from the top-level menu and **Exclude...** from the pull-down menu.
- Select **Direct Links**.
- Click on the atoms which must not be directly linked (i.e. the atoms you want to prevent being bonded to one another).
- Hit **OK.**
- In the example below, atoms 3 and 6 have been specified to form no direct links:

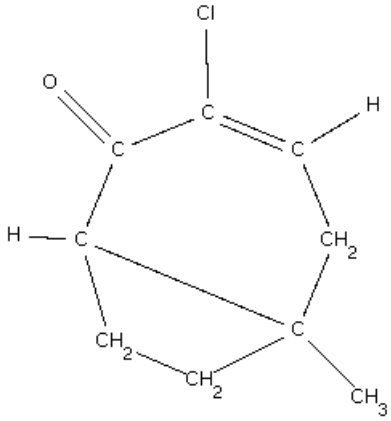




- This query would find:

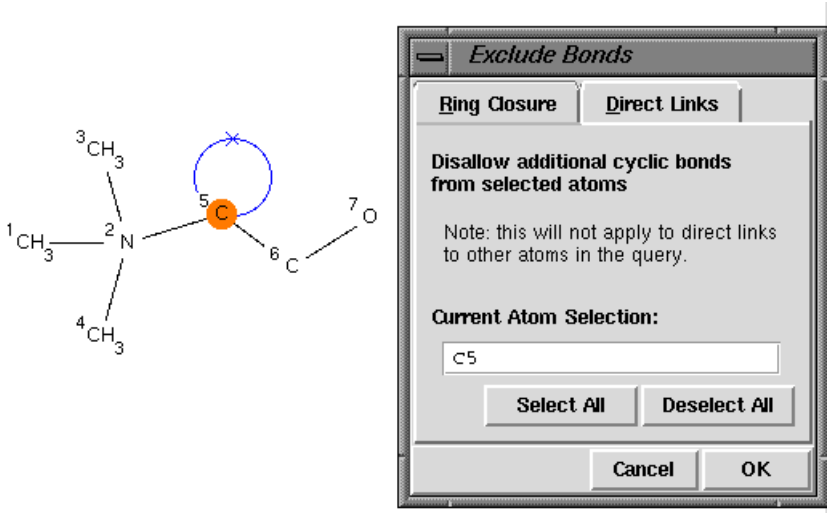


(because the specified atoms are not *directly* bonded to each other) but would not find:

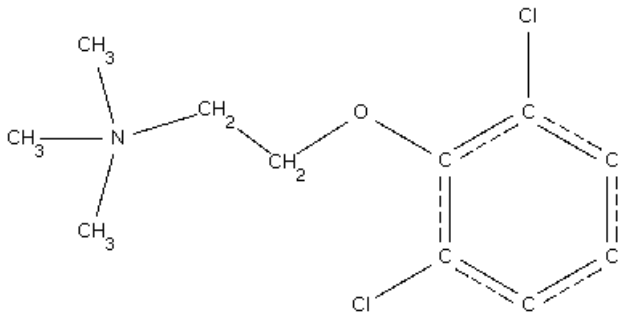


3.8.2 Avoiding Unwanted Cyclic Bonds to Atoms outside the Substructure

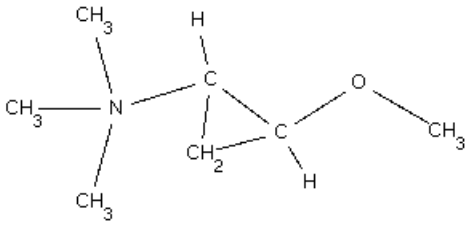
- Select **Bonds** from the top-level menu and **Exclude...** from the pull-down menu.
- Select **Ring Closure**.
- Click on the atoms which must not form cyclic bonds to atoms outside the substructure.
- Hit **OK.**
- In the example below, atom 5 has been selected to form no additional cyclic bonds:



- This query would find:

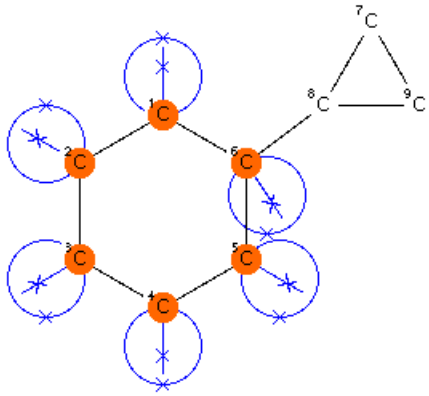


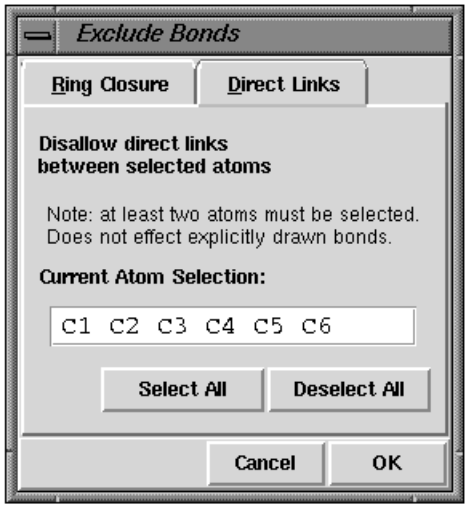
but would not find:



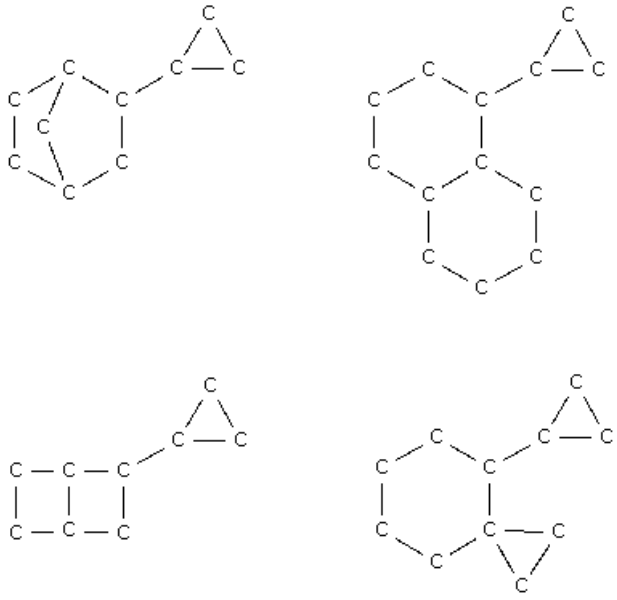
3.8.3 Ensuring that a Ring System has no Additional Bridging or Fusion

- Select **Bonds** from the top-level menu and **Exclude...** from the pull-down menu.
- Select **Ring Closure**.
- Click on all of the atoms in the ring system.
- Select **Direct Links**.
- Click on all of the atoms in the ring system.
- Hit **OK.**
- In the example below, the 6-membered ring system has been *isolated*, i.e. prevented from being part of a larger fused or bridged ring system.





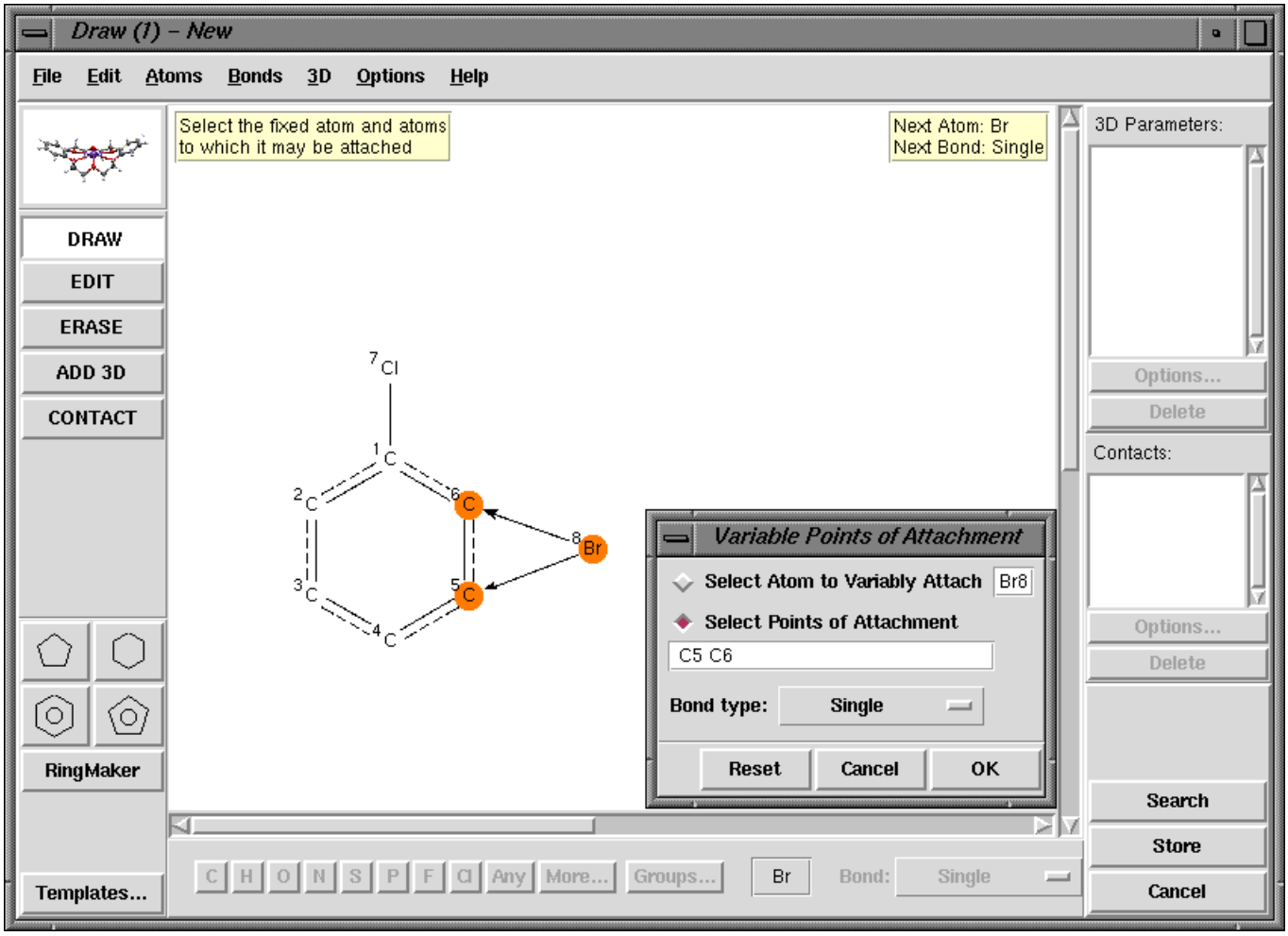
- This query would not find any of the following:



3.9 Specifying Variable Points of Attachment

If you wish to search for both o-bromo-chlorobenzenes and m-bromo-chlorobenzenes (for example) then you can do this in a single search by specifying the o- and m- positions as a variable point of attachment for the bromine atom:

- Having drawn, for example, a chlorobenzene fragment and a Br in the *Draw* window, select **Bonds** from the top-level menu and **Variable Attachment...** from the resulting pull-down menu. This generates a *Variable Points of Attachment* dialogue box:



- Select the atom having a variable point of attachment, i.e. Br, and then select the points of attachment. The selected bond type from the atom to all, in this case two, attachment points will be the same.
- Hit **OK** and start the search.
- This will retrieve hits containing both o-bromo-chlorobenzenes and m-bromo-chlorobenzenes.

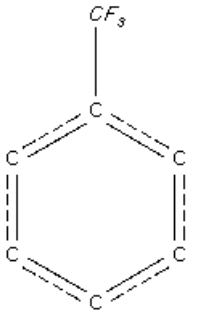
3.10 Finding Molecules Containing Two or More Substructures

If you construct a query in the *Draw* window containing two or more disconnected substructures, you can ensure that these substructures are located in the same chemical molecule by selecting **Atoms** from the top-level menu in the *Draw* window and **All Atoms in Same Molecule** from the resulting pull-down menu.

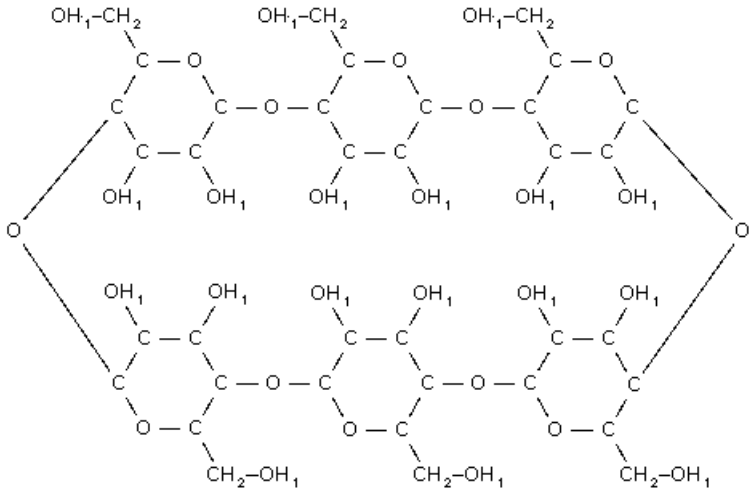
3.11 Using Chemical Groups and Substructure Templates

Substructure drawing can be made easier by using chemical groups and templates, both of which are pre-drawn substructural fragments. Chemical groups are substituents such as *-CF3* which have a

specific point by which they must be attached to the rest of a query substructure, e.g.

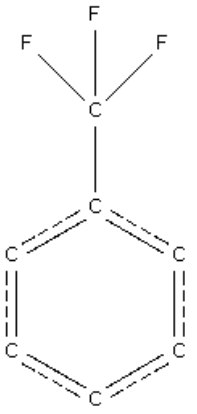


Templates are complete, stand-alone substructural fragments, e.g.



It is possible to:

- Add a chemical group to an existing substructure.
- Replace a terminal atom in an existing substructure by a chemical group.
- Expand a chemical group so that all its atoms and bonds are shown in full, e.g.



- Access and add to the drawing area a standard templates, i.e. one of a set of templates supplied

with ConQuest (see Accessing Standard Substructure Templates, page 71).

- Access and add to the drawing area a customised template, which you have drawn previously and saved for future use (see Saving and Using Customised Templates, page 71).

3.11.1 Adding a Chemical Group to an Existing Atom

This can be done in several ways, including:

- In **DRAW** or **EDIT** mode, right-click on the atom to which the group is to be added, select **Add Group** from the resulting pull-down menu, then pick the group you want from the subsequent menus.
- In **DRAW** or **EDIT** mode, select **Atoms** from the top-level menu, then hit **Add Group** and pick the group you want from the subsequent menus. You will then be asked to pick the atom(s) to which the group is to be added.
- Select the atoms to which you want to add a group. Then hit **Atoms** in the top-level menu, **Add Group** from the resulting pull-down menu, and pick the group you require.
- In **DRAW** mode, change the current element type to a chemical group by hitting the **Groups...** button at the bottom of the window and selecting the required group from the subsequent menus. You can then draw groups in exactly the same way as you would normally draw an atom.
- A complete list of groups can be obtained by selecting the **View...** option, which appears in the menu that is displayed whenever an **Add Group** or **Groups...** button is hit. Groups may be selected from the resulting dialogue box and added to the query substructure.

3.11.2 Changing an Atom to a Chemical Group

This can be done in several ways, including:

- In **DRAW** or **EDIT** mode, right-click on an atom, select **Element** and then **Chemical Groups**, and then pick the desired group from the subsequent pull-down menus.
- In **DRAW** or **EDIT** mode, select **Atoms** from the top-level menu, then hit **Element** followed by **Chemical Groups**, and pick the group you want from the subsequent menus. You will then be asked to pick the atom(s) which are to be replaced by the group.
- Select the atoms which you want to replace. Then hit **Atoms** in the top-level menu, followed by **Element** and **Chemical Groups** in the resulting pull-down menus, and then pick the group you require.
- In **DRAW** mode, change the current element type to a chemical group by hitting the **Groups...** button at the bottom of the window and selecting the required group from the subsequent menus. Then left-click on the atom(s) you wish to replace.
- A complete list of groups can be obtained by selecting the **View...** option, which appears in the menu that is displayed whenever a **Chemical Groups** or **Groups...** button is hit. Groups may be

selected from the resulting dialogue box and added to the query substructure.

- The program will not prevent you from changing a non-terminal atom to a chemical group, but it will never make chemical sense to do so, as a group only has one point of attachment.

3.11.3 Expanding a Chemical Group

- By default, chemical groups are shown as simple, special symbols, e.g. *Et* for ethyl.
- A chemical group may be expanded so that each atom and bond is shown explicitly.
  For example:

  - To expand a chemical group, right-click on the group and select **Expand Group**.
  - Alternatively, select **Atoms** from the top-level menu, **Expand Chemical Groups** from the resulting pull-down menu, then either **Selected** or **All** (if you pick the former, you will be asked to pick the groups to be expanded).

3.11.4 Accessing Standard Substructure Templates

- Several pre-drawn fragments are available to aid substructure drawing.
- To view them, hit the **Templates...** button in the bottom left-hand corner of the *Draw* window, then select **View...** in the resulting pull-down menu. Alternatively, hit **File** in the top-level menu of the *Draw* window, followed by **Import Template** and then **View...**
- The resulting dialogue box not only shows the available templates but allows you to select a template and **Load** it into the drawing area.
- Alternatively, you can add a template to the drawing area by hitting the **Templates...** button in the bottom left-hand corner of the *Draw* window, selecting **List** in the resulting pull-down menu and then choosing the required structural type from the resulting pull-down menu (e.g. **Boron-cages**). The exact template required can then be selected from the next pull-down menu (e.g. **Hexaborane**).
- The same list of templates can also be accessed by hitting **File** in the top-level menu of the *Draw* window, followed by **Import Template** and then **List**.

3.11.5 Saving and Using Customised Templates

- To save the current contents of the white drawing area for future use as a template, pick **File** from the top-level menu followed by **Save Template...** in the resulting pull-down menu. The default file extension is *.cqt*.
- To read in a previously saved substructural template, pick **File** from the top-level menu followed by **Import Template** and **File...** from the resulting pull-down menus.
- To build up a library of customised templates, save them in the templates sub-directory of your search data directory (typically, the search data directory will be a sub-directory in your user area called csds_data).

- Templates saved in this way can be read in by picking **File** from the top-level menu, **Import Template** from the resulting pull-down menu, and **Custom** in the next pull-down menu.

3.12 Substructure Display Conventions

Most of the conventions and symbols used in displaying substructures are obvious. Those that are not include:

- An atom whose symbol begins with the letter *Q (QA, QB*, etc.) has a variable element type. Put the cursor over the atom to get a help message giving further details.
- The symbol *X* indicates an atom that may have **Any** element type.
- The symbol *Z* indicates an atom that can be any element type except hydrogen.
- A superscript beginning with the letter *T* indicates the total coordination number of the atom, e.g. *T4* indicates that the atom must be 4-coordinate.
- H_v next to an atom indicates that the atom is bonded to a variable number of hydrogen atoms. Put the cursor over the atom to get a help message giving further details.
- Chemical groups (accessed via buttons such as **Groups...**, **Add Group** and **Chemical Groups**) are displayed as special symbols, e.g. *Et* for *ethyl*.
- The letter *a* indicates acyclic; *c* indicates cyclic.
- Various conventions are used to display unusual and variable bond types. If in doubt, place the cursor over the bond to get a help message giving further details.
- If an atom is surrounded by a circle, it is close to, or on top of, another atom. Change to **EDIT** mode, select the atom, and move it away (see Moving Atoms, page 73).

3.13 Advanced Drawing Options

3.13.1 Moving Atoms

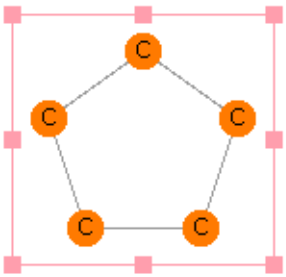
- In **EDIT** mode, select (see Selecting Atoms and Bonds, page 50) the atom(s) to be moved, move the cursor onto one of the selected atoms, press the left-hand mouse button, and move the cursor while keeping the button depressed.
- If two atoms overlap (coloured red) when the mouse button is released, they will be fused.

3.13.2 Rotating Substructures

- It is only possible to rotate complete substructures, not a collection of atoms which form part of a substructure.
- In **EDIT** mode, select (see Selecting Atoms and Bonds, page 50) the substructure to be rotated, move the cursor to a corner of the box surrounding the selected atoms, press the Control key and then the left-hand mouse button. The cursor should change shape to two curved arrows.
- Rotate by moving the cursor, keeping both the mouse button and the Control key depressed.

3.13.3 Resizing Substructures

It is only possible to resize complete substructures, not a collection of atoms which form part of a substructure. In **EDIT** mode, select (see Selecting Atoms and Bonds, page 50) the substructure to be resized, e.g.



- To resize in the horizontal direction only, move the cursor to one of the solid pink squares at the middle of a vertical edge of the fragment's bounding box, press the left-hand mouse button, and move the cursor left or right, keeping the mouse button depressed.
- To resize in the vertical direction only, move the cursor to one of the solid pink squares at the middle of a horizontal edge of the fragment's bounding box, press the left-hand mouse button, and move the cursor up or down, keeping the mouse button depressed.
- To resize equally in both directions, move the cursor to one of the solid pink squares at a corner of the fragment's bounding box, press the left-hand mouse button, and move the cursor, keeping the mouse button depressed.
- To resize unequally in both directions, move the cursor to one of the solid pink squares at a corner of the fragment's bounding box, press the left-hand mouse button, and move the cursor, keeping the mouse button and the Shift key depressed.
- If atoms overlap when the mouse button is released, they will not be merged.

3.13.4 Duplicating Substructures (Copy, Cut and Paste)

To make a copy of all or part of a substructure, select (see Selecting Atoms and Bonds, page 50) the atoms and bonds to be copied and either:

- Click on a blank point in the white area with the right-hand mouse button and select **Copy** or **Cut** (they both make a copy of the fragment; in addition, **Cut** will delete the selected atoms and bonds); or
- hit **Edit** in the top-level menu and **Copy** or **Cut** in the resulting pull-down menu.

To paste a copy of a substructure into the same *Draw* window, or another *Draw* window (in the same ConQuest session), either:

- Click on a blank point in the white area with the right-hand mouse button and select **Paste** (the option will be greyed out if there is nothing available to paste); or
- hit **Edit** in the top-level menu and **Paste** in the pull-down menu.

3.13.5 Changing Default Drawing Options

By default, all bonds drawn from an atom have a fixed length, and other drawing options are also preset. These defaults may be changed using check buttons under the top-level menu item **Options.**

- To change the default length of bonds: ensure **Snap to Grid** is turned on and select **Drawing Options...** to change the value (the *Default bond length* is 60).
- To change the minimum increment that will be made to a bond angle as the cursor is moved when drawing a new bond from an existing atom: ensure **Snap to Grid** is turned on and select **Drawing Options...** to change the value (the *Default bond angle* is 15 degrees).
- To draw bonds freehand, i.e. to the exact position specified by the cursor movement, turn off the **Snap to Grid** check button.
- To change the tolerance for determining how close two atoms must be before they are judged to be overlapping, and how close the cursor must be before it is considered to be *on* an atom: select **Drawing Options...** and change the value (the default *Tolerance* is 15; increase to make the effective size of an atom bigger).

3.14 Pasting in Substructures from Isis/Draw

ConQuest offers the ability to copy and paste substructure queries from ISIS/Draw (MDL Information Systems, Inc.) into either the *Build Queries* area or the *Draw* window (Windows only). In order to use this facility ISIS/Draw must be configured to copy a *MOL* file to the Windows clipboard:

- In ISIS/Draw select **Options** from the top-level menu and **Settings...** from the resulting pull-down menu.
- Hit the *General* tab switch on the **Copy Mol/Rxnfile to Clipboard** check-box.
- Hit **Save** if you wish this change to be applied to subsequent ISIS/DRAW sessions.
- Hit **OK** to close the dialogue box and apply the change to your current ISIS/DRAW session.

Now draw your substructure in ISIS/Draw and copy all or part of the substructure; select **Edit** from the ISIS/Draw top-level menu and **Copy** from the resulting pull-down menu.

To paste a copy of a substructure drawn in ISIS/Draw into the *Build Queries* area:

- Hit **Edit** in the top-level menu of *Build Queries* and **Paste** in the resulting pull-down menu. The pasted substructure will appear listed as a query in the *Build Queries* pane.

To paste a copy of a substructure drawn in ISIS/Draw into the ConQuest *Draw* window, either:

- Click on a blank point in the ConQuest white drawing area with the right-hand mouse button and select **Paste** (the option will be greyed out if there is nothing available to paste); or
- hit **Edit** in the top-level menu of the ConQuest *Draw* window and **Paste** in the resulting pull-down menu.

Not all query features specified in ISIS/Draw can be translated to ConQuest query features. ConQuest attempts to convert the following ISIS query features:

- The majority of atom types including Atom Lists and some Beilstein Atom Generics. If conversion is not possible, the ConQuest Any Atom (**X**) type will be used.
- Bond Type excluding stereochemical features. ConQuest will convert alternating double/single bonds in a six-membered ring to the CSD aromatic bond type. These are the only ring systems where such a conversion is made.
- Zero H-count (no implicit hydrogens).
- Substitution count (number of non-hydrogen attachments with *Any* bond type) if * (as drawn), or zero.
- Ring bond count if * (as drawn) or zero (no ring bonds) if chemically sensible.
- Bond Topology.
- A number of geometric objects such as planes and centroids, and defined geometric parameters, such as distances, angles and dihedrals. More information is available on what is translated and what is not elsewhere (see Saving and Reading Queries, page 19).

It is also possible to copy and paste structures from ChemDraw (CambridgeSoft Corporation) which can also include query features and these are handled in a similar way to ISIS query features.

3.15 Reading MDL MOL Files

MDL MOL files can be read into ConQuest in the same way as a query (.qry) file (see Saving and Reading Queries, page 19).

3.16 Starting Searches and Closing the Draw Window

The *Draw* window can be closed down by:

- Hitting the **Search** button (this starts a search for the substructure(s) in the white drawing area).
- Hitting the **Store** button (this transfers the substructure to the list of queries in *Build Queries*).
- Hitting the **Cancel** button or selecting **File** from the top-level menu and **Close** from the resulting pull-down menu. *Both of these options will discard the current contents of the drawing area.*

3.17 Hints for Substructure Searching

- The key to successful substructure searching is to understand the CSD bond-type conventions (see Bond Type Conventions, page 216), particularly for groups that can be drawn in several valence-bond forms (e.g. Azides (see page 214)). Drawing substructures with incorrect bond types is the usual reason for getting fewer hits than expected.
- If you are unfamiliar with chemical substructure searching, try a few simple searches, e.g. for 6-membered carbocyclic rings, 4-coordinate transition metals.
- Finding exact structures requires complete definition of the target molecule, including H-atoms. If the CSD does not contain the target, then relax the H-atom specification to see if simple derivatives are present.
- In an initial search, do not over-specify the substructure, e.g. in terms of allowed substitution. It is better to get too many hits and then impose tighter chemical constraints. Let the database tell you what it contains!
- If you wish to specify an unsaturated bond but are uncertain as to how it is coded, then use a variable bond type incorporating *double* or *aromatic* or *delocalised double* bonds (see Variable Bond Type, page 246).
- If you are unsure of the bond type which might have been used in the CSD for a particular substructure, use the bond type **Any**, and examine the bond types in the hits generated to formulate a more precise query.
- In many cases, the pattern of bonded atoms is sufficiently specific that close attention to bond types adds little to the search specification. This is particularly true if elements other than C,N or O are specified.

- In tautomeric situations, the crystal structure will contain a specific tautomer. The safe procedure is to formulate the principal tautomers and search for them all.
- Atomic Charge (see page 212) must be associated with specific atoms in a structure. The concept of delocalised charge is not used in the CSD. We do not recommend users to specify charges on transition metal atoms, since the assignment can be rather arbitrary.
- The CSD Element Group Symbols (see page 225) can be used to retrieve families of substructures, e.g. halogeno derivatives, complexes of a specific ligand with any transition metal, etc. You may also define your own element group(s) (see Setting Variable Element Types, page 56).
- The following situations may cause searches to run slowly:
  - Substructures that contain only the most common elements (C,N,O) connected by single or *any* bond types.
  - Substructures that contain one or more locally symmetric units, e.g. t-butyl
  - Searches for large (>12-membered) rings, especially if they are carbocyclic.
- Searches may produce zero hits, or far fewer hits than expected, due to:
  - Chemical errors in the search specification, so check your query carefully.
  - Drawing a resonance hybrid that differs from the CSD convention.
  - Over-specification of the search substructure.

4 TABULATING GEOMETRIES; 3D SUBSTRUCTURE SEARCHES

4.1 Typical Uses for 3D Substructures

- Generating geometry tables for later analysis.
- Locating substructures with specific conformations (torsion angles).
- Locating specific metal coordination geometries, e.g. tetrahedral rather than square planar.
- Pharmacophore searching.

4.2 Basic Guide to Tabulating Geometries

In order to tabulate geometric parameters, they must be defined in the *Draw* window before the search is run, as follows:

- Draw the required substructures (see Fundamentals of Drawing, page 49).
- If necessary, define geometric objects such as planes or centroids (see Defining Geometric Objects, page 80).
- Define the required geometric parameters, e.g. the distance between two atoms (see Defining Specific Geometric Parameters Involving Only Atoms, page 82) or the angle between two planes (see Defining Geometric Parameters Involving Objects, page 86).
- Run the search.
- Transfer the results into Vista or Excel to view the parameters in a spreadsheet (see Transferring Data to Other Programs for Analysis, page 44).

4.3 Basic Guide to 3D Substructure Searching

- Draw the required substructure (see Fundamentals of Drawing, page 49).
- If necessary, define geometric objects such as planes or centroids (see Defining Geometric Objects, page 80).
- Define the required geometric parameters (i.e. the parameters you want to constrain). These may involve just atoms (see Defining Specific Geometric Parameters Involving Only Atoms, page 82) or atoms and objects (see Defining Geometric Parameters Involving Objects, page 86).

- When you define each parameter, constrain it as required, e.g. specify a range of distances for a bond length (see Applying Geometric Constraints, page 88).
- Run the search.

4.4 Geometric Objects

Geometric objects (centroids, planes, etc.) may be defined when drawing a substructure in the *Draw* window. These objects can then be used for computing geometric parameters, e.g. the angle between two planes.

4.4.1 Valid Geometric Objects

Valid objects are:

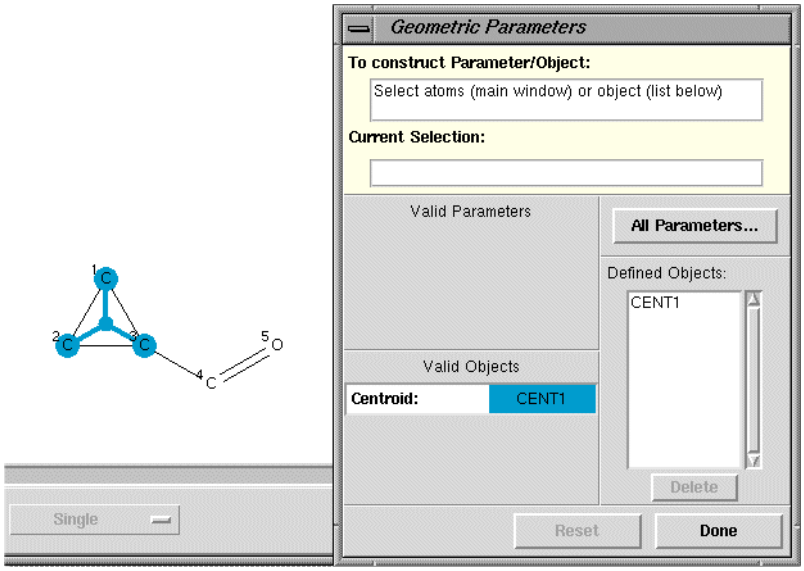
- Centroids (see page 217)
- Dummy Points (see page 225)
- Vectors (see page 246)
- Planes (see page 236)

4.4.2 Defining Geometric Objects

There are two ways to define geometric objects (centroids, vectors, etc.) when drawing a substructure:

Method 1

- Open up the *Geometric Parameters* dialogue box by clicking on the **ADD 3D** mode button or by hitting **3D** in the top-level menu and then **Add Constraints...** in the resulting pull-down menu. The dialogue box can only be opened when there is a substructure in the white drawing area.
- Select the atoms that are needed to calculate the object by clicking on them with the left-hand mouse button (click again on an atom to deselect).
- As the number of selected atoms varies, the dialogue box will list the objects that can meaningfully be defined.
- Hit the appropriate **Define** button in the dialogue box, e.g. next to the word *Centroid* to define a centroid.
- The defined object is listed in the box labelled *Defined Objects*.In the example below, the centroid of the three ring carbon atoms has been defined:



- An extra step is required for the definition of a dummy point: definition of the distance from the second selected atom at which to create the dummy point (this distance can be negative) (see Dummy Points, page 225).

Method 2

- Go to **EDIT** mode and select (see Selecting Atoms and Bonds, page 50) the atoms required for the object.
- Hit **3D** in the top-level menu and then pick the relevant object in the resulting pull-down menu.
- The object will be defined immediately except for Dummy Points (see page 225), where an extra step is required (see above).

Once defined, objects can be used to create geometric parameters, e.g. distances between centroids, or angles between planes (see Defining Geometric Parameters Involving Objects, page 86).

4.4.3 Displaying Geometric Objects in the Draw Window

A defined object may be displayed by clicking on its name in the *Defined Objects* list. This may be found in the *Geometric Parameters* dialogue box (opened by hitting the mode button **ADD 3D**).

4.4.4 Deleting Geometric Objects

An object may be deleted by opening the *Geometric Parameters* dialogue box (click on the **ADD 3D** button), clicking on the object name in the *Defined Objects* list, and then hitting the **Delete** button underneath this list.

4.5 Geometric Parameters

Geometric parameters may be defined when drawing a substructure in the *Draw* window. These parameters can then be listed or tabulated after the search has been run (see Basic Guide to Tabulating Geometries, page 79). They can also be used to set up 3D substructure searches (e.g. a search for a substructure in which a distance has been constrained to a particular range) (see Applying Geometric Constraints, page 88).

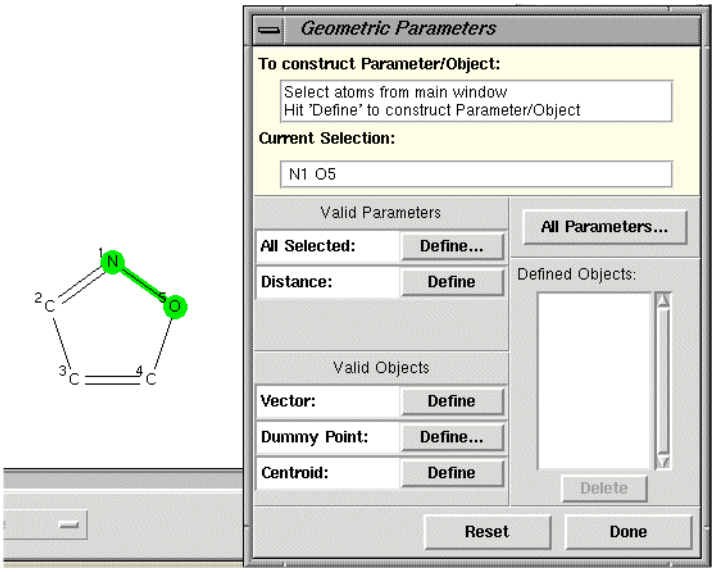
4.5.1 Valid Geometric Parameters

Valid geometric parameters are:

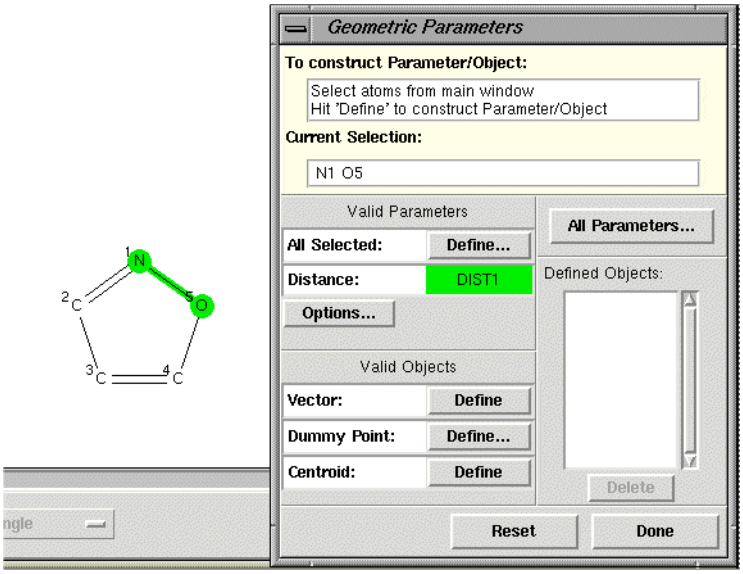
- Atomic Labels (see page 213): although not strictly *geometrical* parameters, it is useful to be able to tabulate the labels of variable atoms (i.e. atoms in the substructure whose element type can vary) (see Setting Variable Element Types, page 56). It is then possible in programs such as Vista (see page 246) to discern differences and trends between different elements at that particular position.
- Distances (see page 225) between atoms and/or objects; the atoms do not need to be bonded to each other.
- Angles (see page 210) between atoms and/or objects; the atoms do not need to be bonded to one another.
- Torsion Angles (see page 242) involving atoms and/or objects; the atoms do not need to be bonded to one another.

4.5.2 Defining Specific Geometric Parameters Involving Only Atoms

- Geometric parameters must be explicitly defined in the *Draw* window in order to be displayed in *View Results* (see Displaying the Values of Defined Geometrical Parameters, page 38), or to be passed to Excel or Vista for tabulation and statistical analysis (see Transferring Data to Other Programs for Analysis, page 44).
- Open up the *Geometric Parameters* dialogue box by clicking on the **ADD 3D** mode button or by hitting **3D** in the top-level menu and then **Add Constraints...** in the resulting pull-down menu. The dialogue box can only be opened when there is a substructure in the white drawing area.
- Select the atoms that are needed to calculate the required parameter by clicking on them with the left-hand mouse button (click again on an atom to deselect).
- As the number of selected atoms varies, the dialogue box will list the parameters that can meaningfully be defined. In the example below, two atoms have been selected so it is possible to define their labels or the distance between them:



- Hit the relevant **Define** button, e.g. next to the word *Distance* to define an interatomic distance. The defined parameter will be listed in the box labelled *3D Parameters* in the top-right hand corner of the *Draw* window. In the example below, the N...O distance has been defined and named *DIST1* by default:



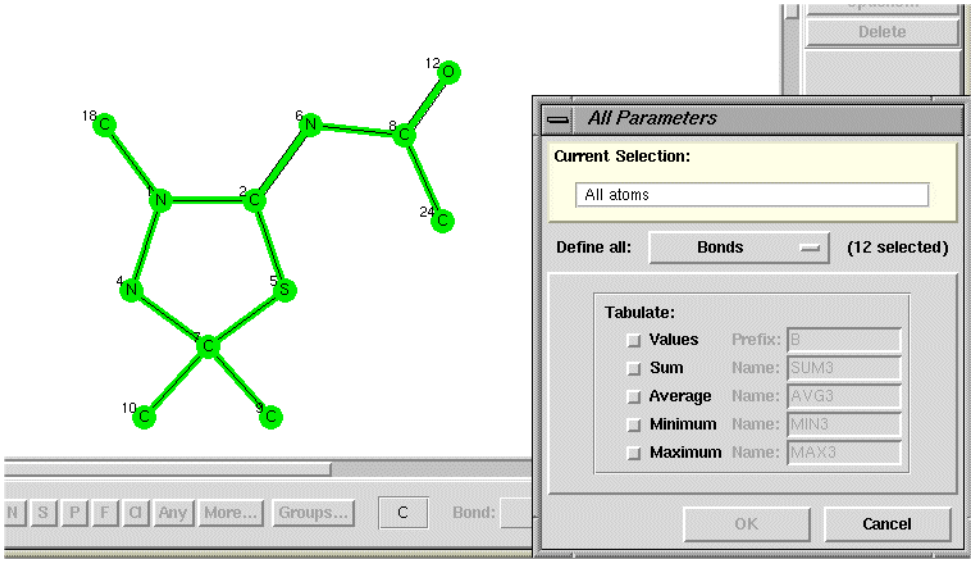
- Once a parameter has been defined, its value can be constrained (see Applying Geometric Constraints, page 88).

4.5.3 Defining All Geometric Parameters Involving Atoms

- All geometric parameters in a substructure, or all parameters described by manually selected atoms within a substructure can be defined.

To define all geometric parameters in a substructure:

- Draw the substructure as normal, then click the ADD 3D button.
- Click on the **All Parameters...** button in the *Geometric Parameters* window.



- The whole substructure will become highlighted and the *All Parameters* window will appear.
- From this window it is possible to define all parameters for atoms, bonds, angles and/or torsions. The number of selected atoms, bonds, angles or torsions in the substructure will be shown to the right of the *Define all:* menu (e.g. *12 selected* in the screenshot above).

To define all geometric parameters described by manually selected atoms in a substructure:

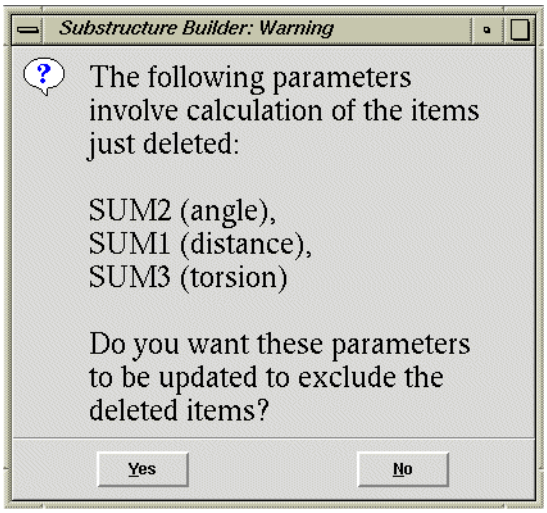
- Draw the substructure as normal, then click on the ADD 3D button.
- Select the atoms that make up the parameters of interest, then click on the **Define...** button, next to the *All Selected:* option.
- This will allow all bond distances, angles and torsions defined by the selected atoms to be tabulated.

- Whichever option is used above, geometric parameters are selected by choosing the relevant geometric parameter from the *Define all:* pull-down menu and clicking on the **Values** radio button. Further data may also be output, depending on the item chosen in the *Define all:* menu:
  - For atoms, the *Atom label*, *Atomic number*, *Atomic radius*, *VdW radius* or *Number of connected atoms* can be defined for output by clicking on the radio button next to the relevant text. Note that the output value for **Number of connected atoms (3D)** reflects the number of bonded atoms in the 3D structure, not the 2D diagram. Thus if H atom positions have not been determined for the structure (e.g. the Os atom in WATWIX) or a particular atom (e.g. the Re atom in VUDSOC10), the bonds to the H atoms will not be included in the *Number of*

connected atoms tabulated value.

- For bonds and angles, values for the following derived parameters can be defined by selecting the radio button next to the appropriate text: *Sum*, *Average*, *Minimum* or *Maximum.*
- For torsions, values for the following derived parameters can be defined by selecting the radio button next to the appropriate text: *Sum*, *Minimum*, *Maximum* or *Use absolute value.*

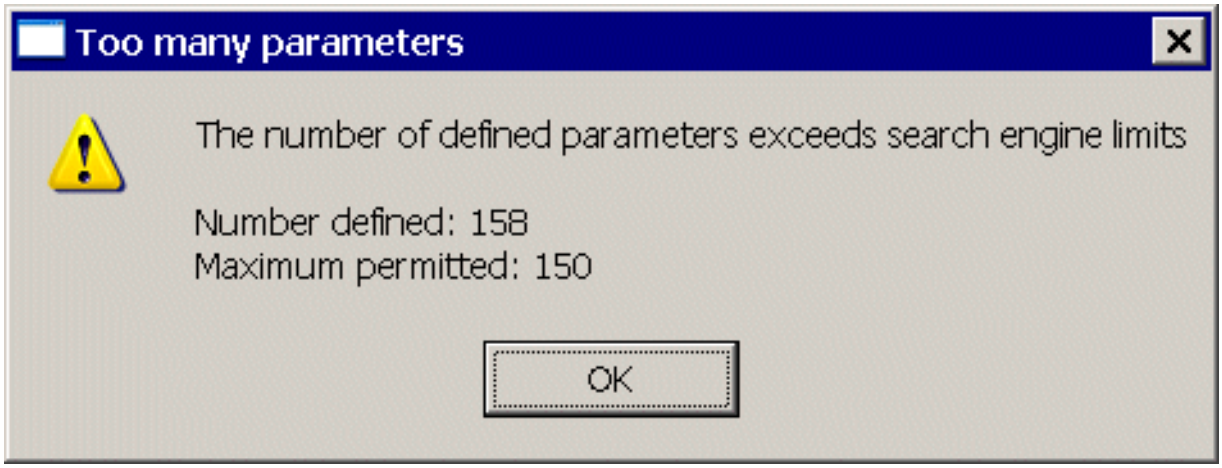
- Constraints can be set for individual bond, angle and torsion measurements, but not for *Sum*, *Average*, *Minimum*, *Maximum* or *Absolute values*.
- If you delete a geometric parameter or bond included in a *Sum* tabulation, ConQuest will issue the following warning:


Substructure Builder: Warning
The following parameters involve calculation of the items just deleted:
SUM2 (angle),
SUM1 (distance),
SUM3 (torsion)
Do you want these parameters to be updated to exclude the deleted items?
Yes
No

- Clicking **Yes** will delete the parameter from the *3D Parameters* list and the *Sum* calculation and the Sum calculation will be updated. Clicking **No** will delete the parameter from the *3D Parameters* list but retain the bond, angle or torsion in the *Sum* calculation.
- It is quite easy to generate a lot of parameters using the *All Parameters* or *All Selected* options. The following features of the *3D Parameters* list in the main *Draw* window may help when manipulating parameters in these situations:
  - If the *Shift* key is pressed whilst clicking on 3D parameters, a continuous range of items will be selected/deselected.
  - If the *Control* key is pressed whilst clicking on 3D parameters, individual items will be selected/deselected.
- If more than one item is selected from the 3D parameter list, the following is possible:
  - If the *Delete* button is hit, all selected items will be deleted.
  - If the *Options...* button is hit, only those selected will be available via the resulting dialogue.

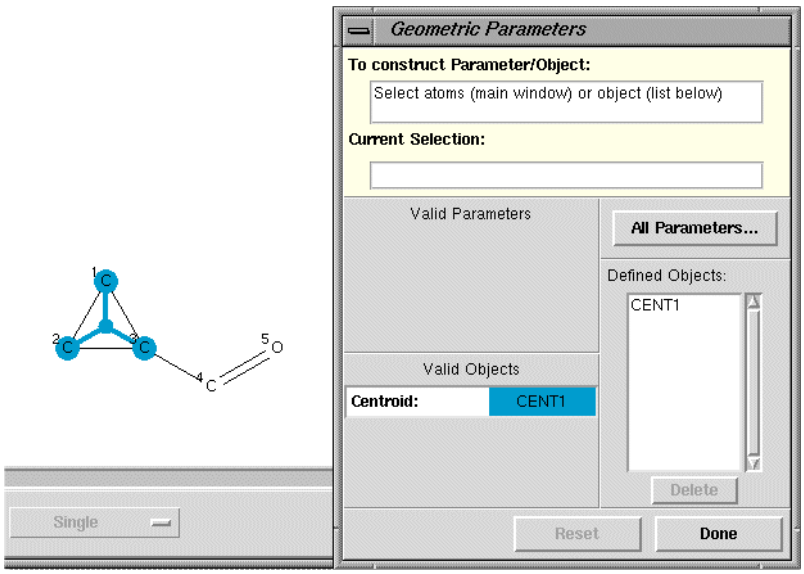
- When using the All Parameters or All Selected it is possible to define large numbers of geometric parameters. A maximum of 150 geometric parameters can be calculated in any one

search: ConQuest will issue a warning if this number is exceeded:

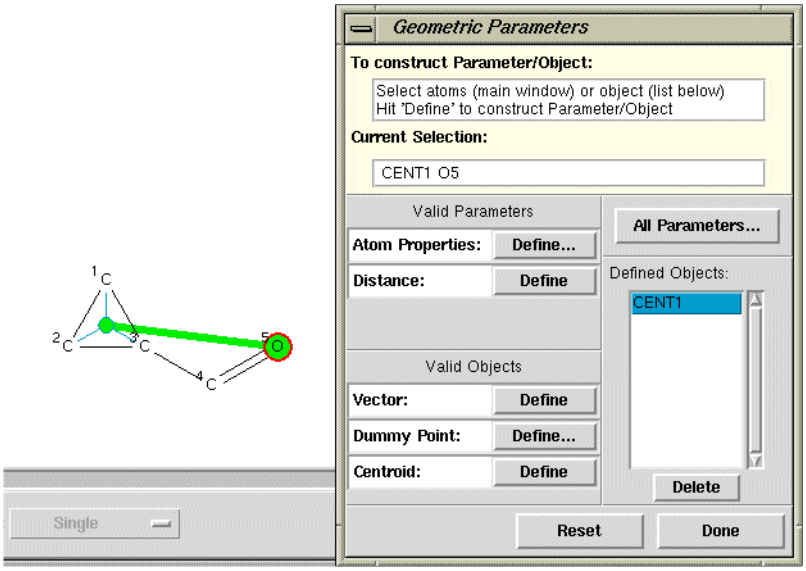


4.5.4 Defining Geometric Parameters Involving Objects

- The procedure is exactly the same as for parameters involving only atoms (see Defining Specific Geometric Parameters Involving Only Atoms, page 82), except that objects are picked from the *Defined Objects* list.
- For example, to specify the distance between an atom and a centroid, first create the centroid (see Defining Geometric Objects, page 80):



- Select the atom by clicking on it, and the centroid by clicking on its object name (*CENT1* in the above example):



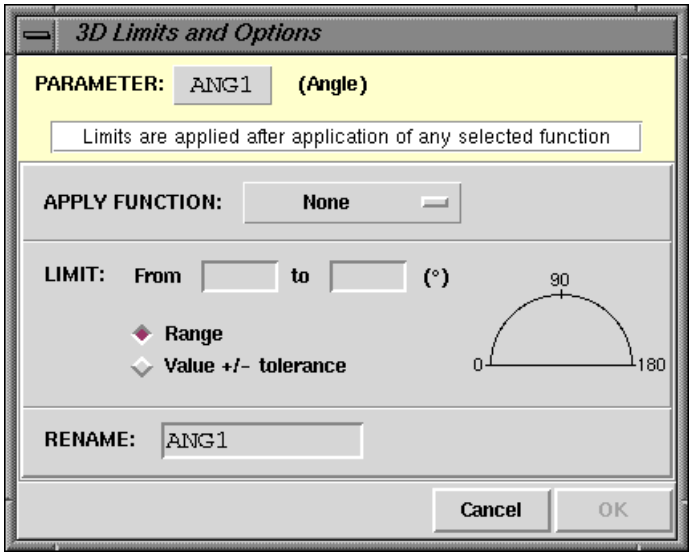
- Then hit the relevant **Define** button (next to the word *Distance* in the above example).
- Once a parameter has been defined, its value can be constrained (see Applying Geometric Constraints, page 88).

4.5.5 Renaming Geometric Parameters

By default, labels are named *LAB1, LAB2*, etc.; distances *DIST1, DIST2,* etc.; angles *ANG1, ANG2,* etc.; torsions *TOR1, TOR2,* etc.

To rename a parameter:

- Select it in the *3D Parameters* list (top right-hand corner of *Draw* window).
- Hit the **Options...** button underneath this list. Alternatively, you can rename a parameter immediately after creating it by hitting the **Options...** button in the *Geometric Parameters* dialogue box (opened by hitting **ADD 3D**).
- Type the new name (maximum 6 alphanumeric characters) into the *RENAME* input box, e.g.



- Hit **OK, Close or Apply**.

4.5.6 Displaying Geometric Parameters in the Draw Window

A defined parameter may be displayed by clicking on its name in the *3D Parameters* list (top-right hand corner of *Draw* window).

4.5.7 Deleting Geometric Parameters

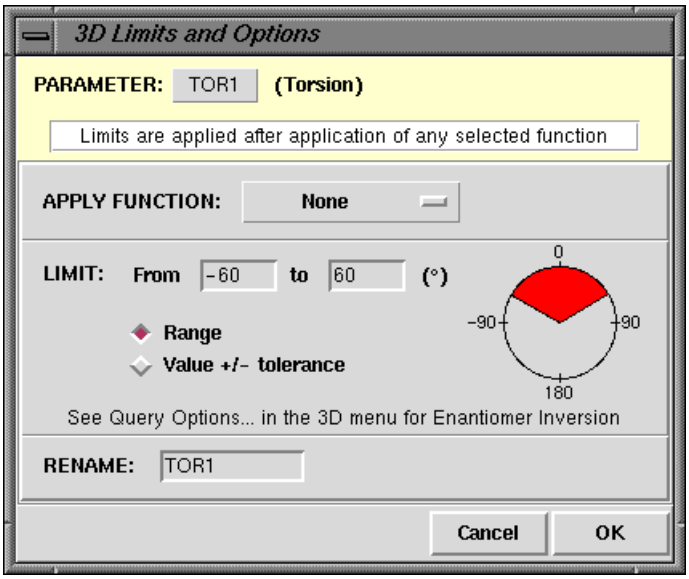
A parameter may be deleted by clicking on its name in the *3D Parameters* list (top-right hand corner of *Draw* window) and then clicking on the **Delete** button underneath this list.

4.6 Applying Geometric Constraints

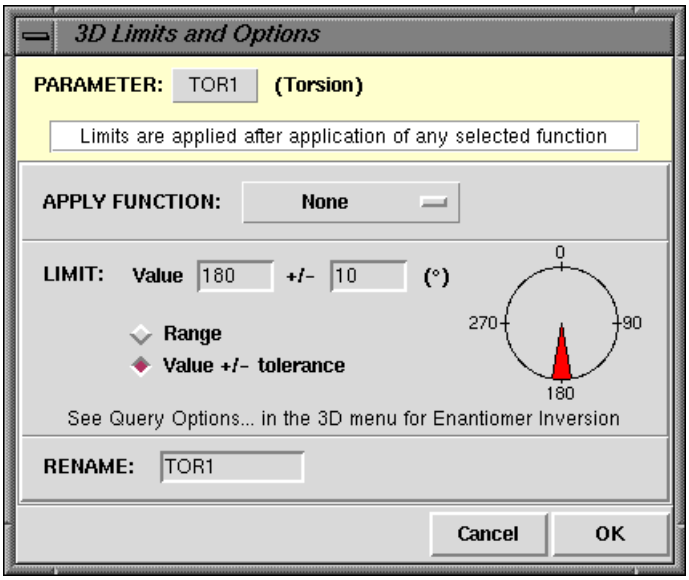
3D substructure searches are performed by defining relevant geometric parameters (see Defining Specific Geometric Parameters Involving Only Atoms, page 82) and constraining their values. To constrain the value of a defined parameter:

- If you have just defined the parameter, so that the *Geometric Parameters* dialogue box is already open, hit the **Options...** button.
- If the *Geometric Parameters* dialogue box is not already open, select the parameter in the *3D Parameters* list (top right-hand corner of the *Draw* window) and hit the **Options...** button underneath this list.
- In the resulting dialogue box, apply, if desired, a function to the parameter (e.g. take the sine of an angle). This can be done by picking from the pull-down menu obtained by pressing the button next to *APPLY FUNCTION*.
- Type the required limits in the *From* and *to* input boxes. In the case of angles, the chosen limits

will be shown graphically in red. In the example below, a torsion angle has been constrained to values between *-60* and *+60* degrees.



- Alternatively, the parameter may be set to a value plus or minus a tolerance. In the example below, the torsion has been constrained to lie within *10* degrees (the tolerance) of *180* (the value):



- A distance constraint may be applied to any pair of atoms, whether bonded or not.
- Hit **OK**, **Apply** or **Close.**

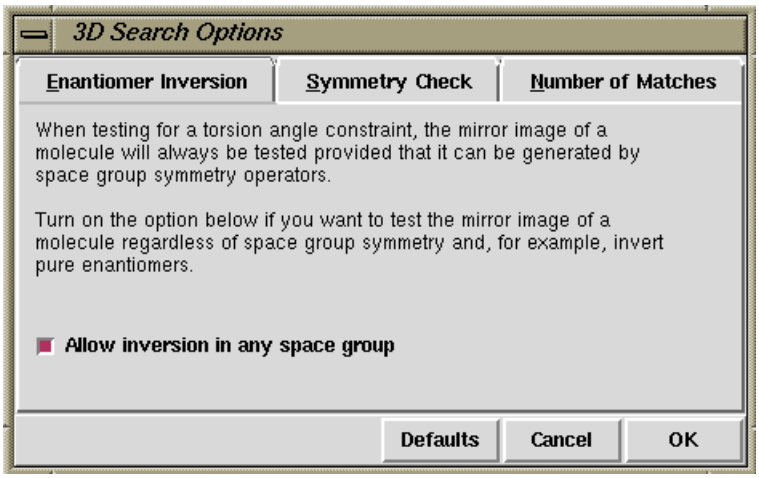
4.6.1 3D Search Options

- Enantiomer Inversion (see page 90)
- Symmetry Check (see Counting the Number of Contacts to an Atom, page 106)
- Number of Matches (see page 90)

Enantiomer Inversion

3D substructure search options are available which allow control of enantiomer inversion when analysing torsion angles:

- In the *Draw* window, once you have defined your geometric parameter of interest select **3D** in the top-level menu and **Query Options...** from the resulting pull-down menu. The *3D Search Options* dialogue box can only be opened when there is a substructure in the white drawing area and you have defined a geometric parameter.
- The *3D Search Options* dialogue box enables you to toggle **Enantiomer Inversion** on and off:


3D Search Options
Enantiomer Inversion
Symmetry Check
Number of Matches
When testing for a torsion angle constraint, the mirror image of a molecule will always be tested provided that it can be generated by space group symmetry operators.
Turn on the option below if you want to test the mirror image of a molecule regardless of space group symmetry and, for example, invert pure enantiomers.
Allow inversion in any space group
Defaults
Cancel
OK

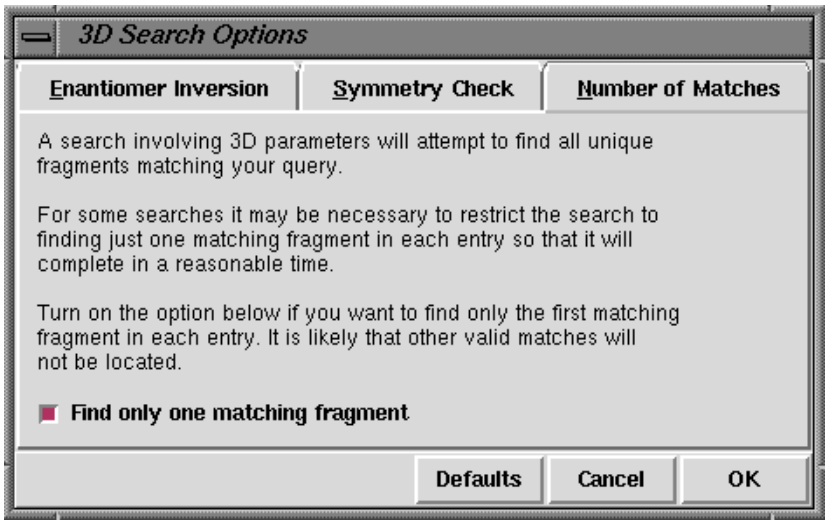
Number of Matches

A substructure search involving 3D parameters will attempt to find all unique fragments matching your query. The need for *exhaustive* fragment location is due to the fact that, although the 2D chemical structure is identical in every case, the 3D structure is almost certainly different. Some crystal fragments may pass the 3D search criteria and be classified as hits; others will not.

In crystal structures there are more problems to consider when carrying out the exhaustive fragment location. We must consider molecules having internal symmetry and where the molecular symmetry coincides with an appropriate symmetry element in the crystal. Also, the asymmetric unit of a crystal structure can contain more than one example of the chemical molecule stored in the CSD.

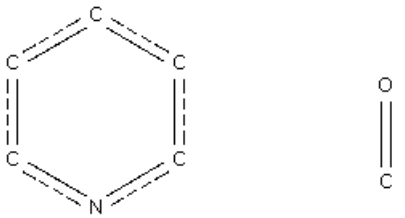
For these reasons, the default and safest mode of operation for the 3D search routines is exhaustive fragment searching, the default setting. However, in some special cases, you may wish to alter this default setting:

- In the *Draw* window, once you have defined your geometric parameter of interest select **3D** in the top-level menu and **Query Options...** from the resulting pull-down menu. The *3D Search Options* dialogue box can only be opened when there is a substructure in the white drawing area and you have defined a geometric parameter.
- The *3D Search Options* dialogue box enables you to select **Find only one matching fragment** from **Number of Matches**:

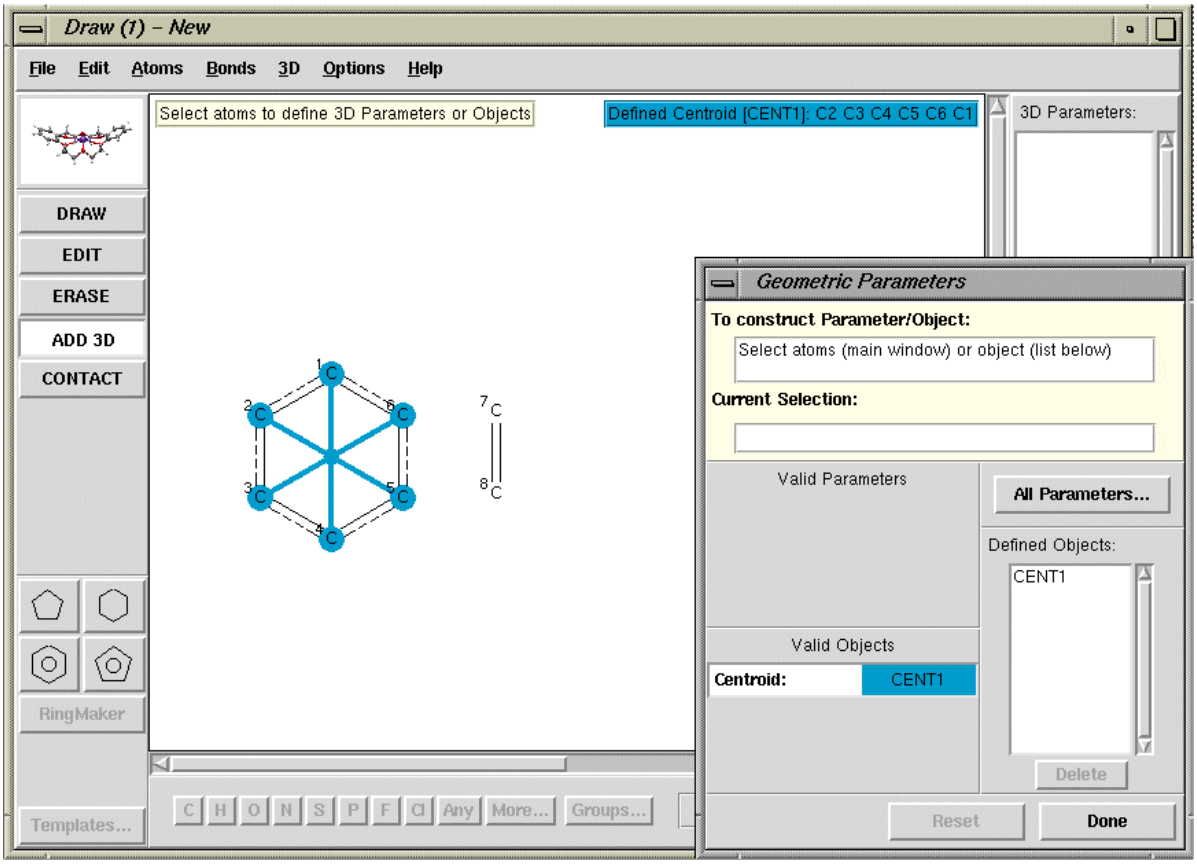

3D Search Options
Enantiomer Inversion
Symmetry Check
Number of Matches
A search involving 3D parameters will attempt to find all unique fragments matching your query.
For some searches it may be necessary to restrict the search to finding just one matching fragment in each entry so that it will complete in a reasonable time.
Turn on the option below if you want to find only the first matching fragment in each entry. It is likely that other valid matches will not be located.
Find only one matching fragment
Defaults
Cancel
OK

4.7 Searching for Molecules Containing Chemical Groups Separated by a Specified Distance; Pharmacophore Searching

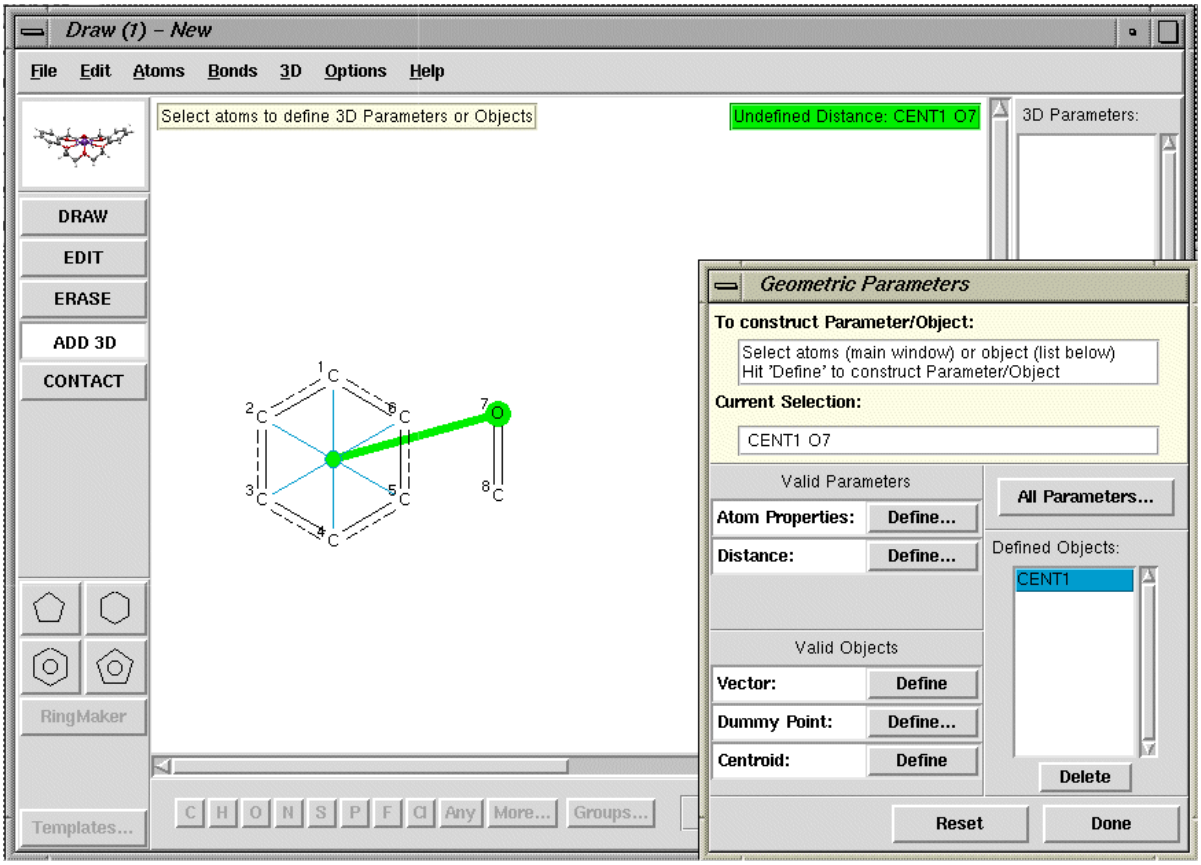
- Draw two (or more) disconnected fragments, e.g.



- If necessary, create any required objects (see Defining Geometric Objects, page 80), e.g. the centroid of a ring:



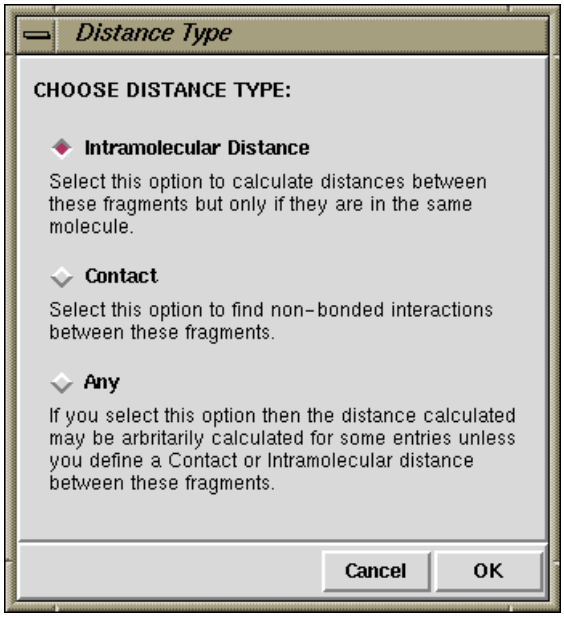
- In **ADD 3D** mode, select the two atoms or objects involved in a distance that you wish to constrain. Atoms are selected by clicking on them, objects are selected by clicking on the object name in the list of *Defined Objects*, e.g.



- Select the **Define** button next to *Distance*. Accept the default (i.e. *Intramolecular Distance*) from the resulting pop-up by hitting **OK**. This will ensure that both fragments are in the same molecule.
- Hit **Options**, type in the required distance limits, and hit **OK**.
- Repeat for any further distance constraints. Other constraints (e.g. angles) may also be applied. Hit **Done** in the *Geometric Parameters* dialogue box when you have finished setting up all the required constraints, then **Search** in the *Draw* window to run the search.

4.8 Types of Distances between Disconnected Substructures

You may be given the following choice of distance types when defining a distance between two disconnected substructures:


Distance Type
CHOOSE DISTANCE TYPE:
Intramolecular Distance
Select this option to calculate distances between these fragments but only if they are in the same molecule.
Contact
Select this option to find non-bonded interactions between these fragments.
Any
If you select this option then the distance calculated may be arbritarily calculated for some entries unless you define a Contact or Intramolecular distance between these fragments.
Cancel
OK

- Select **Intramolecular Distance** if you want the substructures to be in the same molecule. This allows you to find molecules containing two different chemical groups; you will have the option to specify the distance between them (e.g. 4-6Å) or the number of chemical bonds between them (e.g. 3-7 bonds), or both (see Searching for Molecules Containing Chemical Groups Separated by a Specified Distance; Pharmacophore Searching, page 91).
- Select **Contact** if you wish to search for short nonbonded contacts (e.g. hydrogen bonds) between the substructures. You will then be able to choose whether the contacts are intramolecular, intermolecular or both (see NONBONDED CONTACT SEARCHES, page 96).
- Select **Any** if you just want the distance you have selected to be tabulated and intend to define (or have already defined) an intramolecular distance constraint or a nonbonded contact between other atoms in the disconnected substructures.
- Never select **Any** if you do not intend to

  (a) define another distance between the two disconnected substructures, and
  (b) subject that distance to either an intramolecular distance constraint or a nonbonded contact constraint.

  If this rule is not followed, the resulting search will be meaningless, because only distances in the Crystal Chemical Unit (see page 221) will be searched, and the choice of this unit is arbitrary in some space groups.

4.9 Hints for Tabulating Geometries and 3D Substructure Searching

- Geometric parameters are most often defined to generate tables for later analysis using histograms, scattergrams, and statistical tests (see Transferring Data to Other Programs for Analysis, page 44).
- All required geometric parameters must be defined in the *Draw* window when the substructure is

drawn. They cannot be defined after a search has been run.

- Think carefully about the problem being studied to ensure that you have specified geometric parameters that adequately describe that problem. The obvious choice is sometimes not the best.
- Once defined, any geometric parameter can be used as a search constraint by specifying suitable limiting values (see Applying Geometric Constraints, page 88).
- In setting geometric constraints, it is often useful to survey typical values found in the CSD before deciding the limiting values to be used in a subsequent search.
- Chemical environments are usually achiral, so the signs of individual torsion angles may be ignored, and limits on absolute torsion angle values are often satisfactory. Use the *APPLY FUNCTION* option in the *3D Limits and Options* box to take absolute values of torsion angles (see Applying Geometric Constraints, page 88).

5 NONBONDED CONTACT SEARCHES

5.1 Typical Uses of Nonbonded Contact Searching

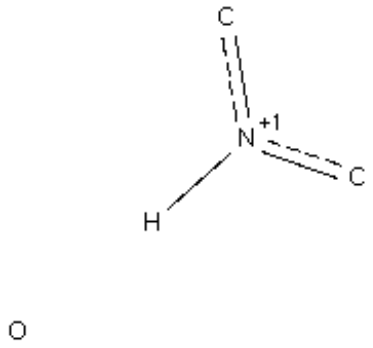
- Finding intermolecular interactions, e.g. hydrogen bonds, halogen...halogen contacts.
- Finding intramolecular non-covalent interactions.
- Generating tables of geometries for nonbonded interactions.

5.2 Nonbonded Contacts between Atoms

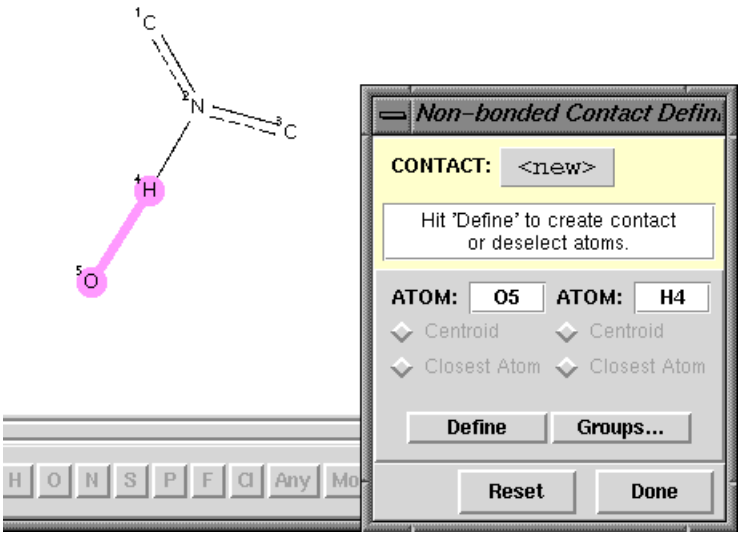
Contacts are defined in terms of distance limits. These can be set by the user (see Specifying Distance Limits for Nonbonded Contacts, page 99).

5.2.1 Searching for Intermolecular Contacts

- In the *Draw* window, draw two separate (i.e. unconnected) substructures (a substructure may consist of a single atom), e.g.



- Either select **Define Nonbonded Contact...** from the top-level **3D** menu or hit the **CONTACT** button located at the left-hand side of the drawing area.
- Pick the two atoms which are to be in contact:



- Hit **Define.**
- By default, the search will find all intermolecular contacts shorter than the sum of the Van der Waals Radii (see page 243) of the atoms involved. To accept this default, hit **Done** and then **Search** in the *Draw* window. (The use of van der Waals radii to specify contact distances is only meaningful for interactions between non-metals.)
- If you are searching for contacts to C-H, N-H or O-H hydrogen atoms, you may wish to normalise the H-atom positions, which corrects for systematic errors in the location of hydrogen atoms by X-ray diffraction. This cannot be done in the *Draw* window. Instead, it must be done in the *Search Setup* dialogue box which is displayed whenever you hit a **Search** button.
- To allow intramolecular contacts (see Searching for Intramolecular Contacts, page 97), or to alter the distance constraint (see Specifying Distance Limits for Nonbonded Contacts, page 99), hit **Edit...**

5.2.2 Searching for Intramolecular Contacts

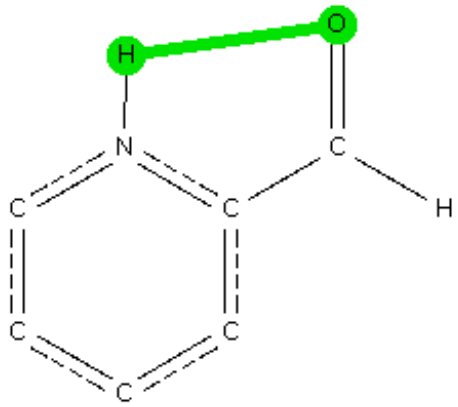
This section describes how to search for intramolecular contacts between two atoms.

- Start by defining the contact as if it were an intermolecular search (see Searching for Intermolecular Contacts, page 96).
- Once the contact has been defined, it can be edited by hitting the **Edit...** button in the *Nonbonded Contact Definition* dialogue box. In the resulting pop-up, use the check-boxes to control whether the search is for **intermolecular** contacts, **intramolecular** contacts or both.
- If **intramolecular** contacts are requested, you may specify the number of chemical bonds

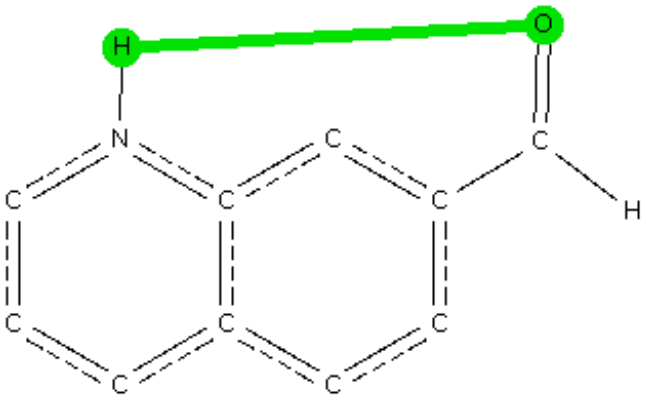
separating the atoms which are in nonbonded contact.

- For example, if the following is specified:

the search would find O...H contacts like:



(where the O and H atoms are separated by four chemical bonds) but not:



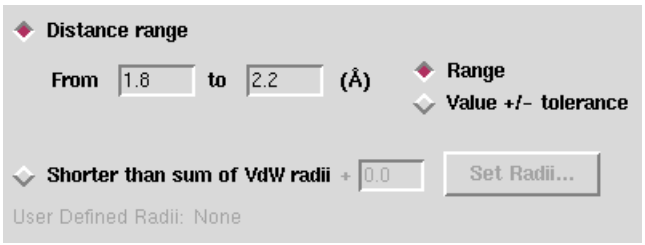
(where the O and H atoms are separated by six bonds).

5.2.3 Specifying Distance Limits for Nonbonded Contacts

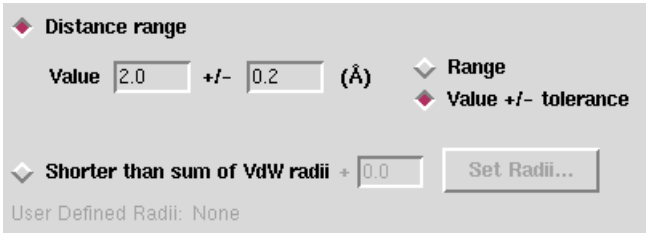
Once a contact between two atoms has been defined (see Searching for Intermolecular Contacts, page 96), it can be edited by hitting the **Edit...** button in the *Nonbonded Contact Definition* dialogue box. In the resulting pop-up, use the check-boxes to control whether the distance between the interacting atoms is to be specified relative to the Van der Waals Radii (see page 243) of the atoms.

Three choices are possible:

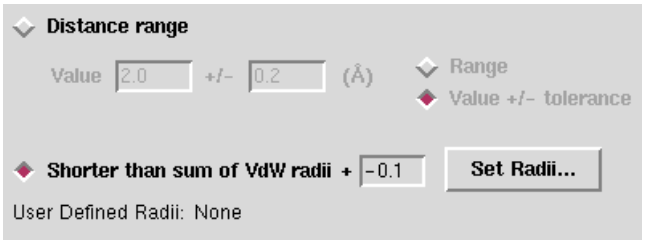
- If **Distance Range** and **Range** are both switched on, you may type in the minimum and maximum distances between the interacting atoms. The example below specifies a distance range of 1.8 - 2.2 Å:



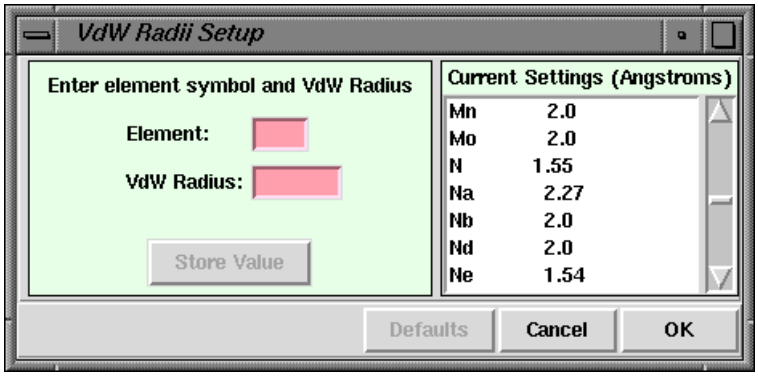
- If **Distance range** and **Value +/- tolerance** are both switched on, you may type in a target value with a tolerance allowed either side. The example below specifies that the minimum interatomic distance is 2.0 - 0.2 = 1.8Å, and the maximum is 2.0 + 0.2 = 2.2Å:



- If **Shorter than sum of VdW radii** is switched on, the van der Waals radii of the interacting atoms are added together, to give a quantity V. The search will then find all contacts shorter than V+x, where x is the value typed in the input box. The example below will find all contacts shorter than V - 0.1Å, where V is the sum of the van der Waals radii of the interacting atoms. Thus, for a contact between O (vdw radius = 1.52Å) and H (vdw radius = 1.20Å), contacts shorter than 1.52 + 1.20 - 0.1 = 2.62Å would be found:



Setting of van der Waals radii is possible by selecting **Set Radii...** which generates the *VdW Radii Setup* dialogue box, showing the current values and providing boxes for you to enter the element symbol and *new* van der Waals radius (this is particularly important when searching for contacts involving metal atoms where CCDC default radii are not well-defined):

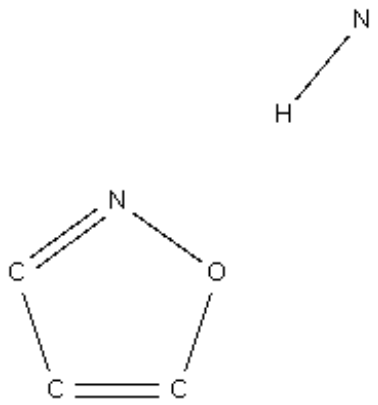


5.3 Nonbonded Contacts between Chemical Groups

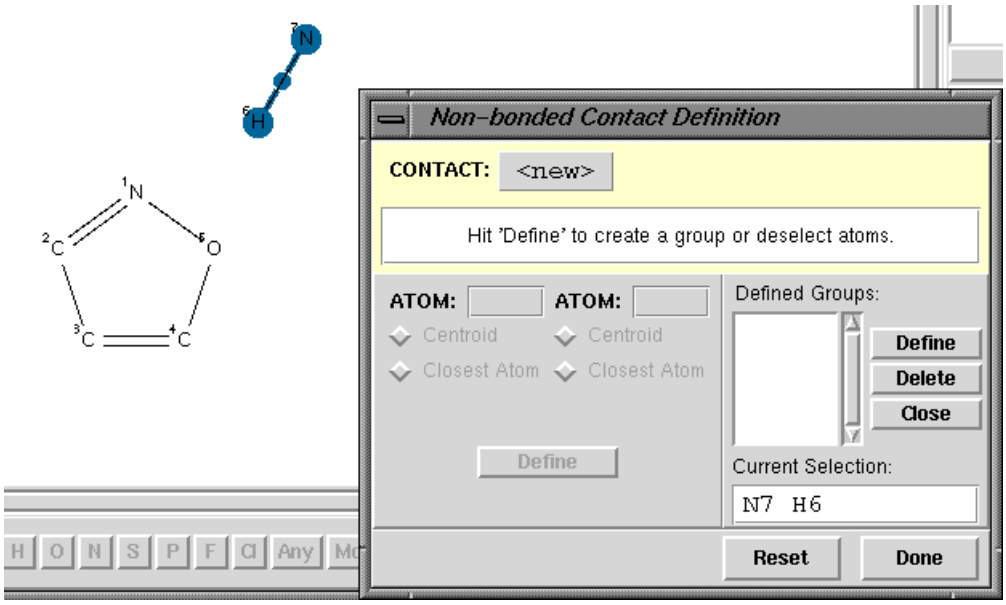
- It is possible to search for nonbonded contacts between chemical groups, each of which may contain any number of atoms.
- The two groups are considered to be in contact if the closest pair of atoms (one from one group, one from the other) are within a user-specified distance or if the distance between the centroids of the two groups is in a specified distance range.

5.3.1 Defining Groups for Use in Nonbonded Searching

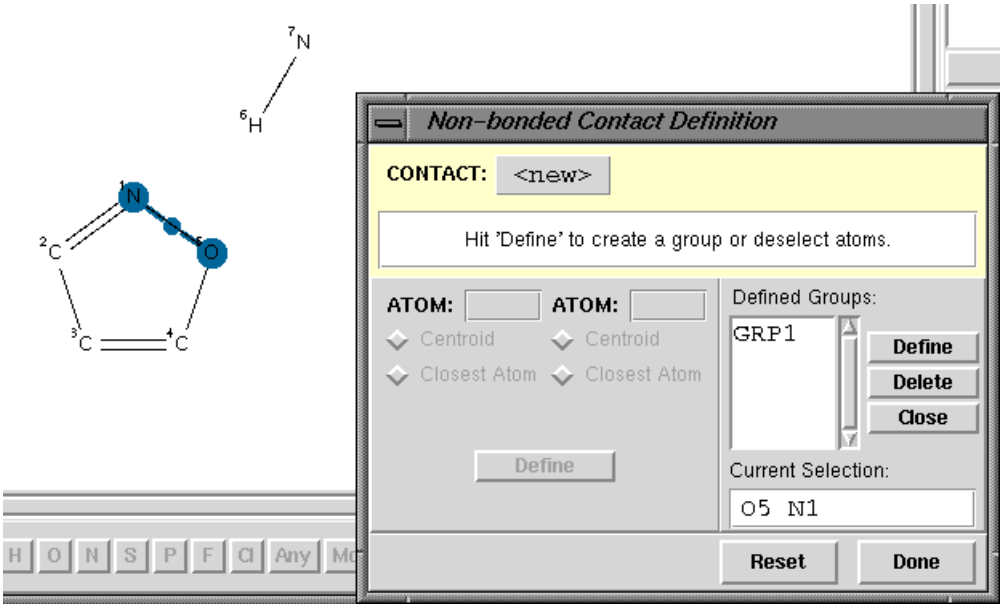
- Draw two unconnected substructures in the *Draw* window (see Layout of the Draw Window, page 47):



- Either select **Define Nonbonded Contact...** from the top-level **3D** menu or hit the **CONTACT** button located at the left-hand side of the drawing area.
- Hit **Groups...** in the resulting pop-up window.
- Pick all the atoms that you wish to be included in the first group:



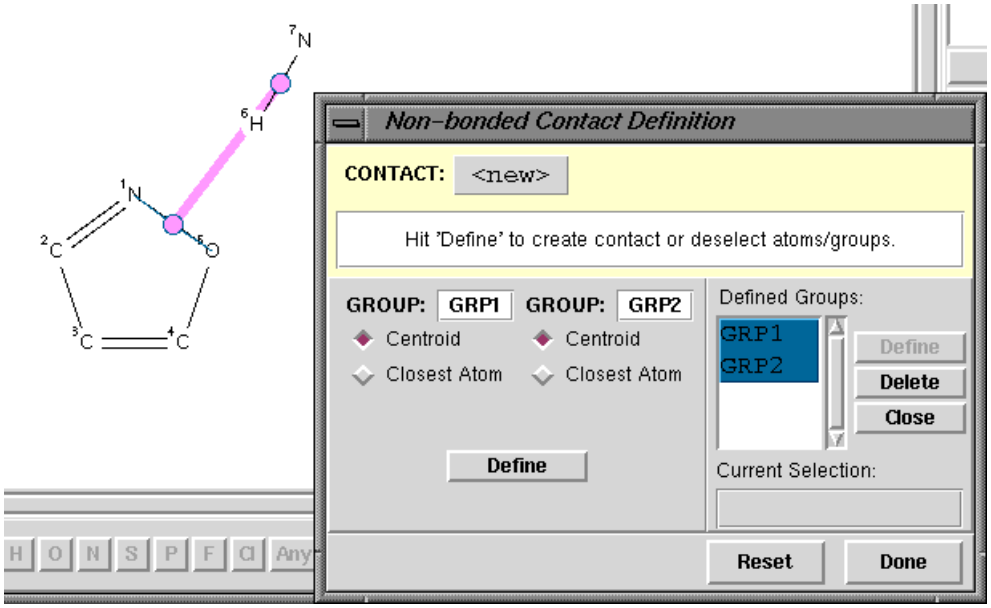
- Hit **Define**. This defines a group, called *GRP1*, which contains the atoms you have selected (the N and H atoms in the above case).
- Pick all the required atoms in the second group:



- Hit **Define**. This defines a second group, *GRP2*, containing, in the above example, the N and O atoms of the ring.
- Select *GRP1* and *GRP2* from the *Nonbonded Contact Definition* pop-up.
- Specify whether the search is to be based on the distance between the centroids of the two groups (see Searching for Chemical Groups whose Centroids are Separated by a Specified Distance, page 102), or the distance between their closest atoms (see Searching for Chemical Groups whose Closest Atoms are Separated by a Specified Distance, page 104).

5.3.2 Searching for Chemical Groups whose Centroids are Separated by a Specified Distance

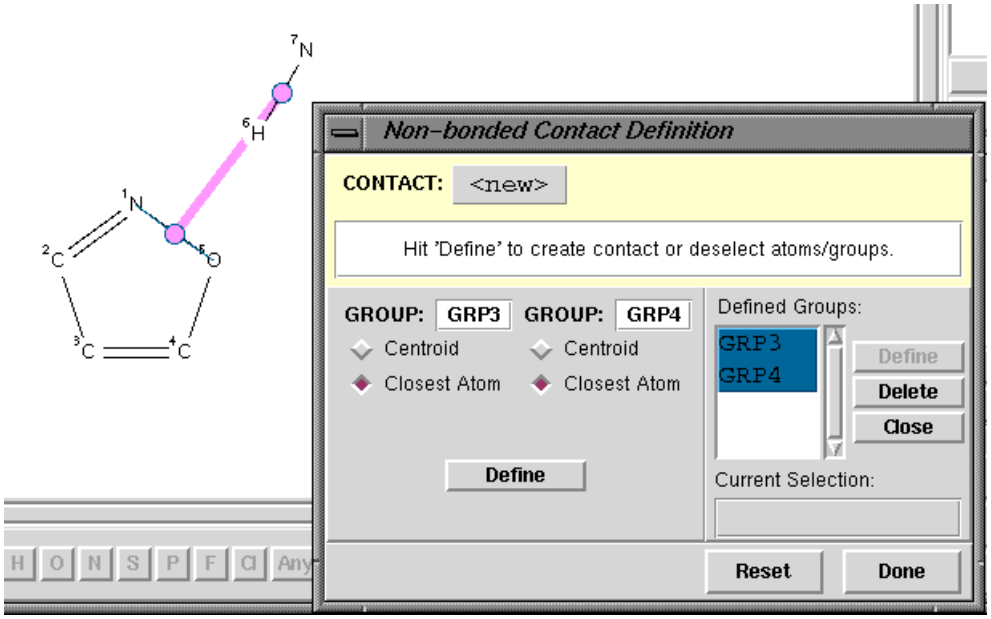
- **Define** the two groups (see Defining Groups for Use in Nonbonded Searching, page 101).
- Select both of the groups in the *Defined Groups* list box and ensure that the **Centroid** check-buttons are switched on:


Non-bonded Contact Definition
CONTACT:
<new>
Hit 'Define' to create contact or deselect atoms/groups.
GROUP:
GRP1
GROUP:
GRP2
Centroid
Closest Atom
Centroid
Closest Atom
Define
Defined Groups:
GRP1
GRP2
Define
Delete
Close
Current Selection:
Reset
Done
H
O
N
S
P
F
Cl
Any

- Click on the **Define** button and specify the distance required between the two group centroids. In the above example, the centroids will be (a) that of the N and H atom and (b) the centroid of the N and O atoms of the ring.
- By default, the search will find all intermolecular contacts in the specified distance range. To accept this default, hit **Done** and then **Search.**
- To allow intramolecular contacts, hit **Edit...**

5.3.3 Searching for Chemical Groups whose Closest Atoms are Separated by a Specified Distance

- **Define** the two groups (see Defining Groups for Use in Nonbonded Searching, page 101).
- Select both the groups in the *Defined Groups* list box and ensure that the **Closest Atom** check-buttons are switched on:

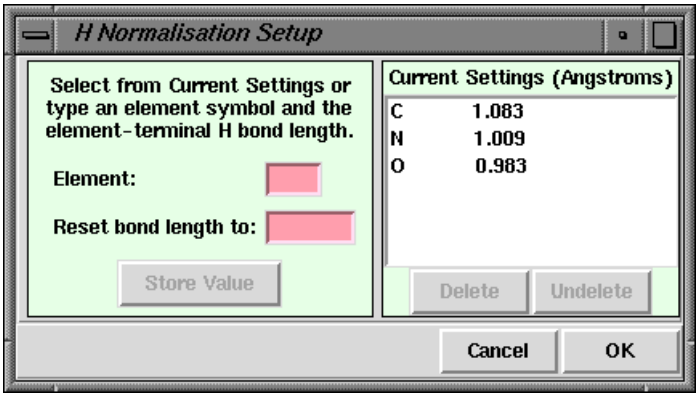


- Click on the **Define** button.
- This specifies a search that will find all pairs of groups (in the above example, an N-H group and the N-O atoms of an isoxazole ring) whose closest atoms lie in a specified distance range.
- By default, the search will find intermolecular contacts shorter than the sum of the Van der Waals Radii (see page 243) of the atoms involved. To accept this default, hit **Done** and then **Search**. Distances based on van der Waals radii are only meaningful for contacts between non-metals.
- If you are searching for contacts to C-H, N-H or O-H groups, you may wish to normalise the H-atom positions, which corrects for systematic errors in the location of hydrogen atoms by X-ray diffraction. This cannot be done in the *Draw* window. Instead, it must be done in the *Search Setup* dialogue box which is displayed whenever you hit a **Search** button.

- To allow intramolecular contacts (see Searching for Intramolecular Contacts, page 97), or to alter the distance constraint (see Specifying Distance Limits for Nonbonded Contacts, page 99), hit **Edit...**

5.4 Correcting the Positions of Hydrogen Atoms

- Hydrogen-atom positions determined by X-ray diffraction are unreliable because hydrogen has a low scattering power for X-rays. In addition, the lengths of covalent bonds involving hydrogen are almost invariably too short when measured by X-ray diffraction, because the centroid of the hydrogen-atom electron density is displaced towards the heavy atom to which the hydrogen is bonded.
- It is desirable to correct for these errors when searching for nonbonded contacts to hydrogen atoms, such as hydrogen bonds. Otherwise, parameters such as hydrogen-bond distances will be systematically incorrect.
- The errors can be corrected by normalising the hydrogen atom positions. This involves moving the hydrogen atom along the covalent bond vector (X-H) so as to make the X-H distance equal to the average neutron diffraction value (hydrogen atoms are accurately located by neutron diffraction).
- Normalisation of hydrogen atom positions is available for C-H, O-H and N-H bonds (the H atom will be moved along the bond vector to make the bond length equal to the average neutron diffraction value (see Hydrogen Atom Location in Crystal Structure Analyses, page 228). To switch it on, set up the nonbonded contact search as usual, and hit a **Search** button. This opens the *Search Setup* dialogue box. Hit the tab labelled *Advanced Options* and switch on the **Normalise terminal H positions** check-box. Then start the search.
- Once you have switched on the **Normalise terminal H positions** option, it will remain on for the rest of the ConQuest session (i.e. for any subsequent searches you may perform) unless you turn it off again.
- Customisation of the values for C-H, N-H and O-H H-normalisation is available from the *Search Setup* dialogue box. Hit the tab labelled *Advanced Options* and switch on the check-box labelled **User Defined**, then hit **Edit...** this generates a *H Normalisation Setup* dialogue box into which you can type the relevant values. It is also possible to specify values to normalise the position of H atoms bonded to other elements.

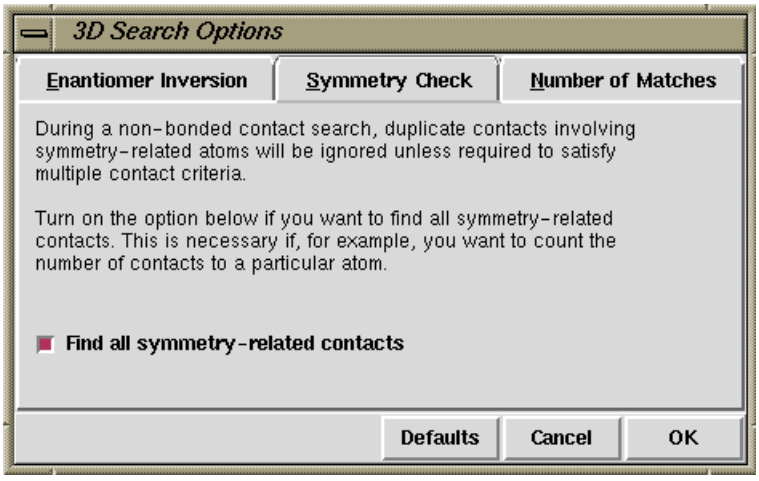


- Normalised hydrogen atom positions may be written out when entries are exported in mol2 and sd formats (see Saving Individual Structures and Lists of Structures, page 41).

5.5 Counting the Number of Contacts to an Atom

In locating nonbonded contacts, the normal ConQuest search mechanism will locate unique, symmetry-independent contacts, e.g. O-H....N1; if a symmetry-related contact O-H....N1* exists, it will, by default, be ignored. However, if you need to know the precise number of contacts to an atom, the **Symmetry Check** should be turned *off*:

- In the *Draw* window, once you have defined your geometric parameter of interest select **3D** in the top-level menu and **Query Options...** from the resulting pull-down menu. The *3D Search Options* dialogue box can only be opened when there is a substructure in the white drawing area and you have defined a geometric parameter.
- The *3D Search Options* dialogue box enables you to switch **Symmetry Check** on and off:


3D Search Options
Enantiomer Inversion
Symmetry Check
Number of Matches
During a non-bonded contact search, duplicate contacts involving symmetry-related atoms will be ignored unless required to satisfy multiple contact criteria.
Turn on the option below if you want to find all symmetry-related contacts. This is necessary if, for example, you want to count the number of contacts to a particular atom.
Find all symmetry-related contacts
Defaults
Cancel
OK

5.6 Hints for Nonbonded Contact Searching

- Searches for intermolecular interactions take longer than those for purely covalent substructures.

Searches for patterns involving two or more non-covalent interactions can be very time-consuming.

- A typical search consists of two (occasionally more) covalent substructures and the non-covalent interactions required between them. The non-covalent interactions must be defined geometrically. For example, if you are searching for O-H...O hydrogen bonds, you must specify the maximum permissible O...O or H...O distance.
- The most common geometrical constraints are maximum and minimum distances, e.g. D...A or H...A distances for hydrogen bonds D-H...A (D = donor, A= acceptor).
- In choosing non-covalent distance constraints, it can help to carry out a preliminary search using quite generous limits (e.g. sum of van der Waals radii plus 0.5Å). This establishes the range of values in the CSD.
- It is only sensible to specify distance constraints with respect to van der Waals radii if you are searching for contacts between the common non-metallic elements unless you choose to set your own van der Waals radii.
- H-atoms are not always present in CSD entries, particularly in older entries. You may need to use D....A distance constraints instead of H...A constraints in H-bonding studies.
- H-atoms are poorly defined by X-ray diffraction: the direction of D-H bonds is usually satisfactory, but the D-H valence-bond distances are shortened. You can correct for this by normalising H-atom positions.
- Other geometric constraints can be used to further define the non-covalent interaction, especially if the interaction is directional. The angle D-H...A in a hydrogen bond prefers to be linear, so some constraints on this angle may be appropriate, particularly for stronger hydrogen bonds. It is wise to examine typical values in the CSD before setting such geometrical constraints.
- Scattergrams from IsoStar, distributed as part of the CSD System, are valuable for assessing search constraints.
- Sometimes it is better to combine searches. For example, all H-bonds involving N or O acceptors and N-H or O-H donors can be found with a generic search for X-H...X, where X is assigned the variable element type *N or O* (see Setting Variable Element Types, page 56). The labels of the X atoms actually found in any interaction can be output by ConQuest, so that the different types of H-bonds can be separated out after searching, e.g. in Vista (see page 246).
- Searches for common interactions, e.g. O-H....O(sp^2) hydrogen bonds, can generate very large numbers of hits. The number of hits can be reduced by either (a) defining the chemical environment of the groups more exactly (keto-O, carboxylate-O, etc.), or (b) lowering the R-factor (see page 239) acceptance limit (e.g. R <= 0.05).

6 PEPTIDE SEARCHES

6.1 Typical Uses of Peptide Searching

- Finding peptide structures containing particular amino acid sequences, e.g. ALA-PRO-PHE.
- Searching for sequences in which one or more of the residues is variable, e.g. ALA-XXX-PHE, where XXX = GLY, PHE, or VAL.
- Finding peptide structures whose sequences contain unusual or chemically modified amino acids (e.g. norvaline, ornithine, sarcosine) or residues linked by non-peptidic bonds.

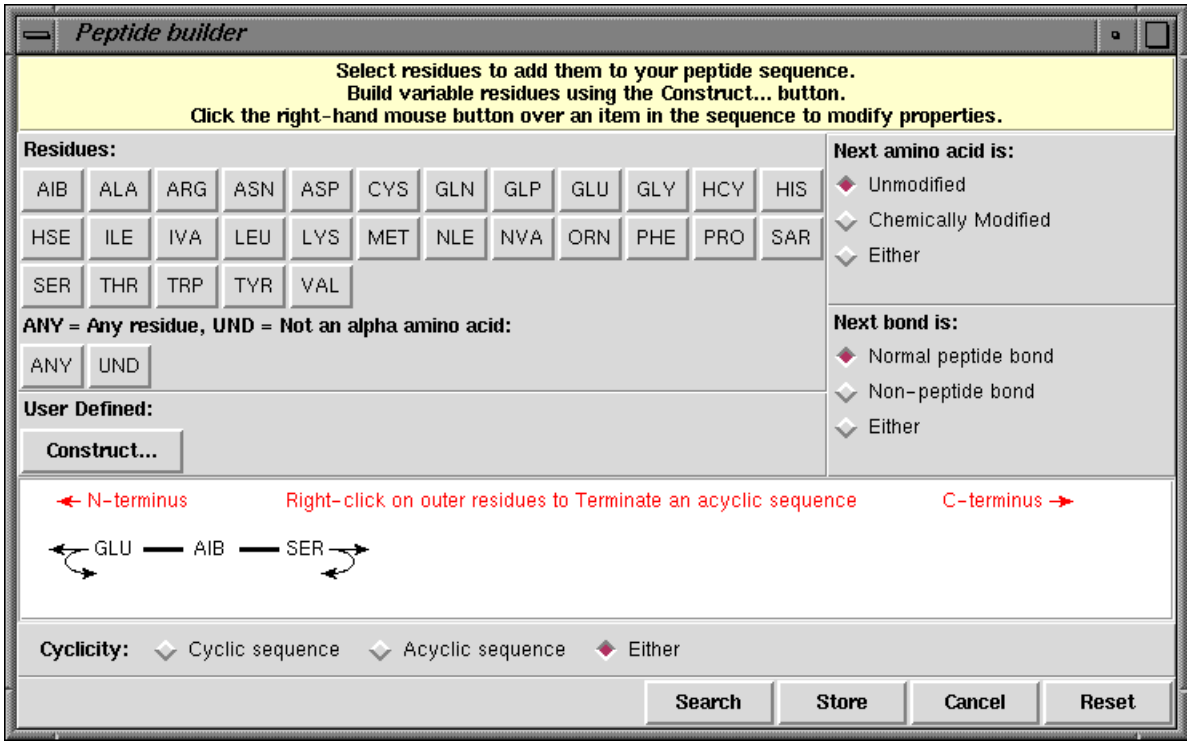
6.2 Using the Peptide Dialogue Box

Open the dialogue box by hitting **Peptide** in *Build Queries*. You can:

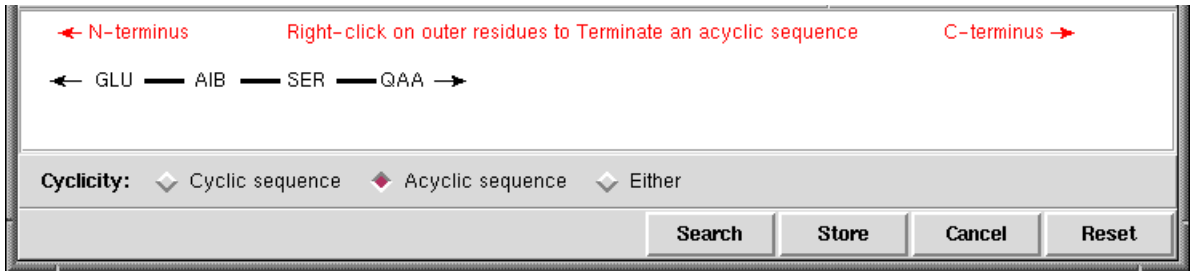
- Search for peptides containing a specific sequence of amino acids.
- Search for a generified sequence, i.e. a sequence in which one or more of the residues can be variable.
- Search for a sequence containing a non-amino-acid unit, or an amino acid that is chemically modified in some way.

6.2.1 Searching for a Specific Peptide Sequence

- Open the *Peptide builder* dialogue box by hitting **Peptide** in *Build Queries*.
- Specify the required peptide sequence by clicking on amino acid buttons:



- Some of the amino acids are non-standard; leave the cursor over a button to get a display of the full chemical name.
- By default, both cyclic and acyclic peptide sequences will be found. This can be altered by using the buttons in the *Cyclicity* section of the dialogue box. For example, only acyclic peptides will be found with the following setting:



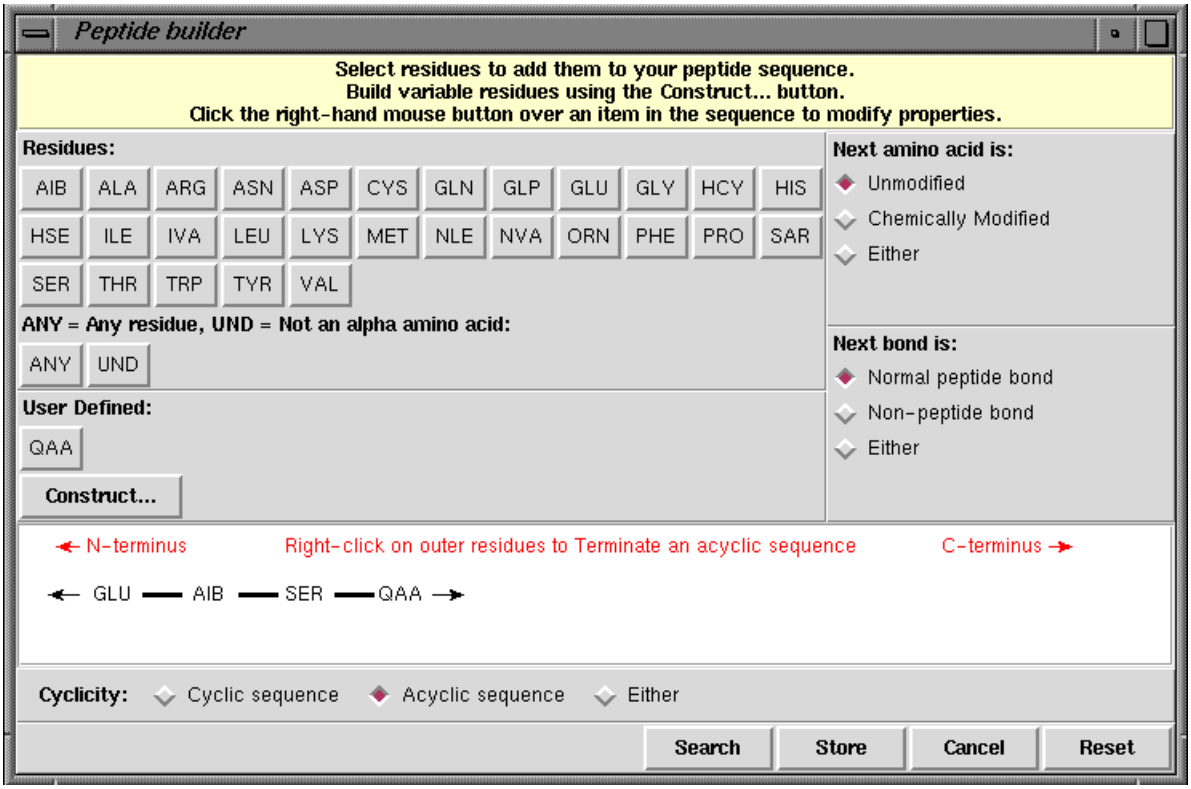
- The buttons underneath *Next bond is* can be used to specify whether the next residue to be added is to be linked by a **Normal peptide bond**, a **Non-peptide bond**, or **Either**. A peptide bond is defined as a link between the C=O group bound to the alpha-carbon of one of the residues and the NH group bound to the alpha-carbon of the next residue.
- A residue can be deleted from a sequence by right-clicking on it and selecting **Delete** from the resulting pull-down menu.

- A residue in a sequence can be changed by right-clicking on it, selecting **Replace by** in the resulting pull-down menu, and then **Residues** in the next menu. This produces a series of menus from which the desired amino acid can be selected. Alternatively, hitting **Replace by** followed by **Other** or **User defined** allows selection of a variable amino acid type.
- An end residue in a sequence can be specified as terminal by right-clicking on it.
- A bond in a sequence can be changed to a different type (i.e. peptide, non-peptide or either) by right-clicking on it, selecting **Modify** in the resulting pull-down menu, and then picking the required type.
- Variable residues can be defined and included in the sequence, if desired.

6.2.2 Specifying a Variable Amino Acid Residue

Searches may be performed for *generified sequences*, where one or more of the residues is variable.

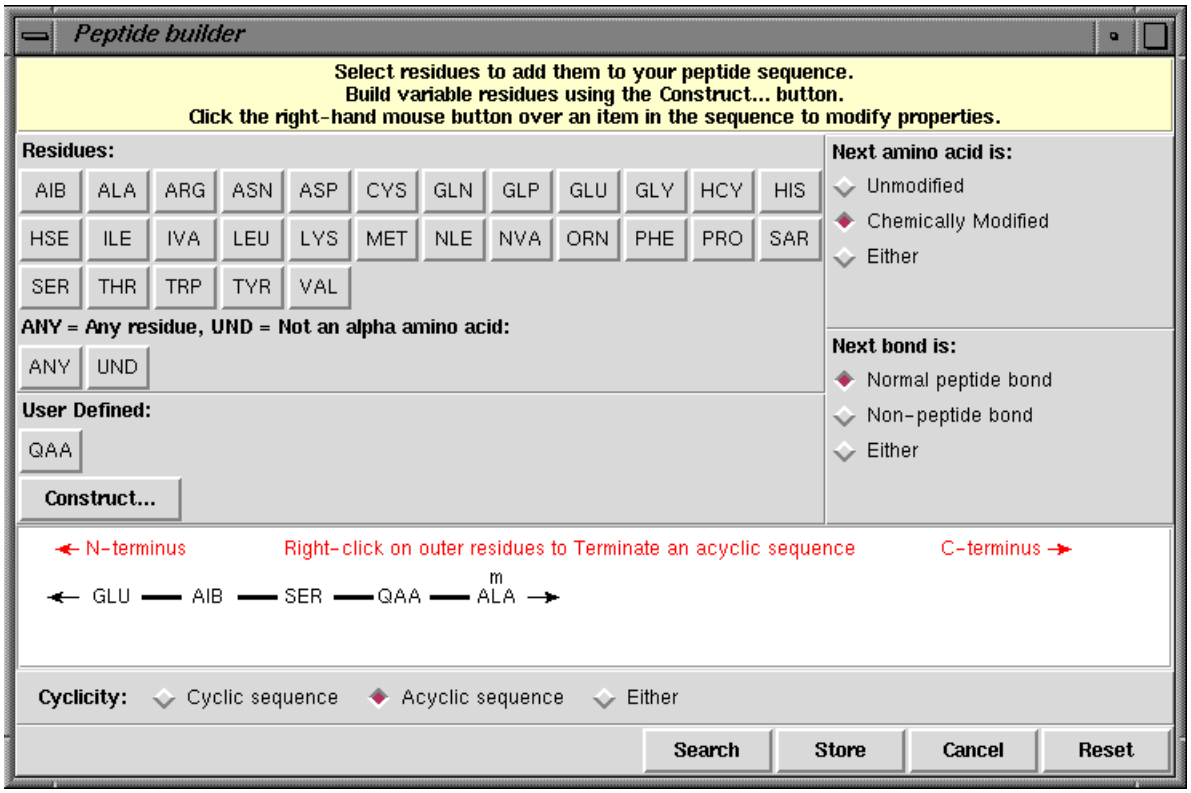
- Use the **ANY** button to add a residue which will match any amino acid. For example, GLU-ANY-SER will find all of GLU-AIB-SER, GLU-ALA-SER, GLU-ARG-SER, etc.
- To define a variable residue, hit the **Construct...** button, select the required amino acids from the resulting *Build user defined residues* dialogue box and hit **Apply**. This will add a new button labelled **QAA** to the *Peptide builder* dialogue box. Clicking on this button (after closing *Build user defined residues* dialogue box) will then add a QAA residue to the sequence, e.g.



- QAA in this sequence will match any of the amino acids that you selected in the *Build user defined residues* dialogue box.
- Repeated use of the **Construct...** button will create further buttons labelled **QAB**, **QAC**, etc., each representing a variable residue type according to your specification.
- Use the **Match any except selection** button in the *Build user defined residues* dialogue box to specify a residue that can be anything except the amino acids you have clicked on.

6.2.3 Specifying a Chemically Modified Amino Acid or a Non-Amino-Acid

- Use the **UND** button to add a peptide residue to the sequence that will match anything except an alpha amino acid. For example, ALA-UND-ARG will find compounds containing an alanine and an arginine residue separated by something that is not an alpha amino acid.
- To allow an amino acid to be chemically modified in some unspecified way (e.g. acetylated) switch on the button labelled **Either** in the section of the dialogue box labelled *Next amino acid is;* to force an amino acid to be chemically modified switch on **Chemically Modified.** Then hit one of the residue buttons. For example, to add a chemically-modified CYS to the sequence, switch on **Chemically Modified**, then hit **CYS**.
- Alternatively, right-click on a residue in the sequence you have built, select **Modifications** from the resulting pull-down menu, then **Unmodified**, **Chemically Modified,** or **Either**, as desired.
- Modified amino acids are indicated by an **m** on top of the peptide residue, e.g



6.3 Hints for Peptide Searching

- If you include a chemically modified residue in a sequence, you cannot specify the type of chemical modification. For example, selecting **Chemically Modified** and **TYR** will match with any modified form of tyrosine, but not with tyrosine itself.
- A peptide search will find any peptide containing the sequence you have specified. For example, ALA-SER would match with GLY-ALA-SER, ALA-SER-ASN, ARG-ALA-SER-PHE, etc., as well as with ALA-SER itself.

7 AUTHOR AND JOURNAL SEARCHES

7.1 Typical Uses of Author and Journal Searching

- Finding publications by specific author(s).
- Finding structures mentioned in particular papers.
- Reviewing publications in specific journal(s) and year(s).

7.2 Using the Author/Journal Dialogue Box

Open the dialogue box by hitting **Author/Journal** in *Build Queries*. You can search for:

- Authors' names (see Authors' Names, page 113).
- Journal name (see Journal Name, page 114).
- Journal volume (see Journal Volume, page 114).
- Journal page number (see Journal Page, page 115).
- Year of publication (see Journal Year, page 115).
- CCDC number or *refcode* (see CCDC number (deposition number), page 115)

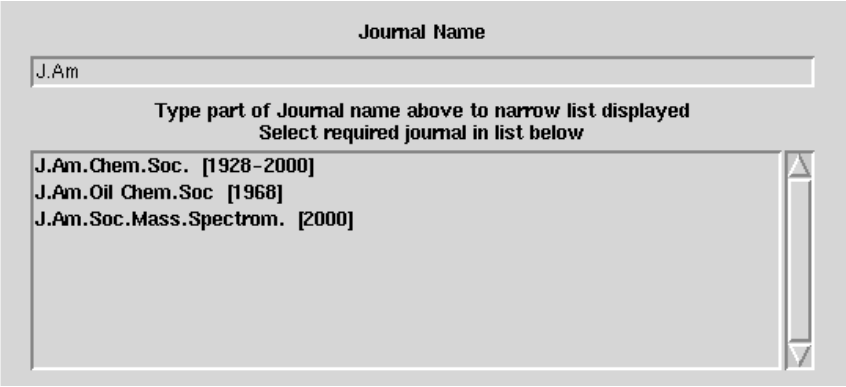
7.2.1 Authors' Names

- Type the required surname into the input box.
- Optionally, authors' initials may be specified, e.g. *F.H.Allen*. Each initial must be followed by a full stop, with no spaces between initials, or between initials and surname. All initials must match exactly, e.g. *F.H.Allen* would not find *F.Allen.*
- If the **Exact surname** check button is turned off, a search for *Allen* would also find *Allenby*, *Allenford*, etc. If the box is turned on, only the exact name *Allen* would be found.
- The **Exact surname** option cannot be used if initials are specified.
- Only one name is allowed per box: use the **New Box** button if you wish to search for structures published by two or more authors. The example below would find structures in papers on which both *Raithby* and *Allen* are authors:

- To find structures published by either *Raithby* or *Allen*, create and store two separate queries, one for Author = *Raithby*, one for Author = *Allen*, and combine them (see Complex Query Combinations, page 21).

7.2.2 Journal Name

- The journal name must be abbreviated *exactly as it is stored in the CSD*. It is therefore safer to select it from the list box rather than type it in.
- To find the journal in the list, either scroll down using the scroll bar, or type the first letter or two of the journal name into the input box. As the name is typed, the list in the box is narrowed down:



- Pick the required journal with a single mouse click.
- If you type in part of a journal name and the input box turns pink (and/or the list box becomes empty), this means that there is no journal in the CSD beginning with the letters you have typed. Backspace to empty the input box completely and browse down the list of journals to see if you can spot the one you want.
- Some journals were only published between certain years (see APPENDIX B: CHANGES IN JOURNAL NAMES, page 248). The years in parentheses next to a particular journal name in the list box refer to the years for which entries are included in the CSD for that particular journal title.
- Searching for the partial name *Acta Cryst* would find all the different sections of *Acta Crystallographica*.

7.2.3 Journal Volume

- Type the required volume number into the input box.
- It is not possible to specify a range.
- Volume numbers sometimes contain letters.
- Some journals do not use volume numbers.

7.2.4 Journal Page

- Type the starting page number into the input box.
- It is not possible to specify a range.
- Page numbers occasionally contain letters.

7.2.5 Journal Year

- Select the required option (**during, after, between**, etc.).
- Type a four-digit year, e.g. *1998*, into the input box.
- If **between** is selected, two years must be specified, and the second must be later than the first, e.g. *1978-1980* is valid, *1978-1977* is not.

7.2.6 CCDC number (deposition number)

- Type the entire deposition number e.g. 243822, 1267/1231.
- It is not possible to search for IUCr deposition codes (e.g. A20110) using this method.
- Recent CCDC numbers (200000 or greater) are shown as hyperlinks; if the hyperlink is selected, ConQuest will attempt to open the *CCDC CIF Repository Request Form* in a Web browser. This form is also available from the CCDC website: *www.ccdc.cam.ac.uk/products/csd/request/*.

7.3 Hints for Author and Journal Searching

- It is easy to miss hits unless you understand the CSD conventions for Authors' Names (see page 213).
- Select journal names from the list box rather than typing them in, since you need to use the same abbreviations as are used in the CSD.
- Some journals have changed their names over time, often by dividing into two or more sections (see APPENDIX B: CHANGES IN JOURNAL NAMES, page 248).
- Some volume numbers contain letters, e.g. *C471*. However, in other cases the alphabetic part of a volume number is absorbed into the journal name, e.g. *Acta Cryst., Section B*. You can usually work out which convention has been used by looking at the way in which the journal name is written in the list box.
- Some journals do not have volume numbers, most notably the journals of the UK Royal Society of Chemistry.
- Journal pages are normally numeric, but can occasionally contain letters, e.g. *C-1431*.
- The year must contain four digits, e.g. *1987*.

8 COMPOUND NAME AND CHEMICAL CLASS SEARCHES

8.1 Typical Uses of Compound Name or Class Searching

- Using names or partial names as a quick way of finding examples of a certain type of structure, e.g. thiazole derivatives (but search for *thiazol*, not *thiazole*).
- Using common names to locate drugs and natural products, e.g. *morphine*.
- Finding certain types of compounds which are difficult to find by substructure searching, e.g. *rotaxanes, catenanes*.
- Retrieving all entries that belong to a specific class or classes of compounds e.g. *Nucleotides and nucleosides*, *Steroids*.

8.2 Using the Compound Name and Chemical Class Dialogue Box

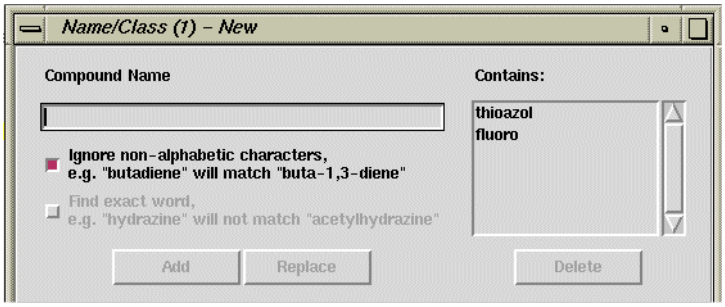
The Name/Class dialogue box is split into two parts:

- *Compound Name* (see Compound Name Searching, page 116)
- *Chemical Class* (see Using the Chemical Class Dialogue Box, page 117)

It is also possible to combine *Compound Name* and *Chemical Class* searches (see Combining Compound Name and Chemical Class Queries, page 118).

8.2.1 Compound Name Searching

- Open the dialogue box by hitting **Name/Class** in *Build Queries*.
- Type the required compound name, or partial compound name, into the input box.
- If the tick box **Ignore non-alphabetic character** is ticked, the search will find structures even if the specified name is interrupted by numbers, commas, hyphens, etc.; e.g. *thiazolyl* would find a structure with the name Azobenzene-N-(4,5-bis(methoxycarbonyl)thiazol-2-yl)imide.
- If the tick box **Find exact word only** is ticked, the search will only find structures in which the compound name matches an exact word; e.g. *hydrazine* will not match *acetylhydrazine*. This type of search cannot be used in conjunction with the **ignore non-alphabetic character** option.
- When finished, click **Add** beneath the **Find exact word only** tick box. The query text will be added to the *Contains:* box.
- To remove an entry from the *Contains:* box, select the relevant word and click on the **Delete** button.
- To edit an entry in the *Contains:* box, select the relevant word, make the amendment then click on the **Replace** button. The query text will be updated.
- To search for more than one name or partial name, type the new query text into the *Compound Name* box and click **Add**. The example below would find all structures whose compound names contain both *thiazol* and *fluoro,* e.g. 4,4,5,5-tetrafluoro-1,3,2-dithiazolidine:

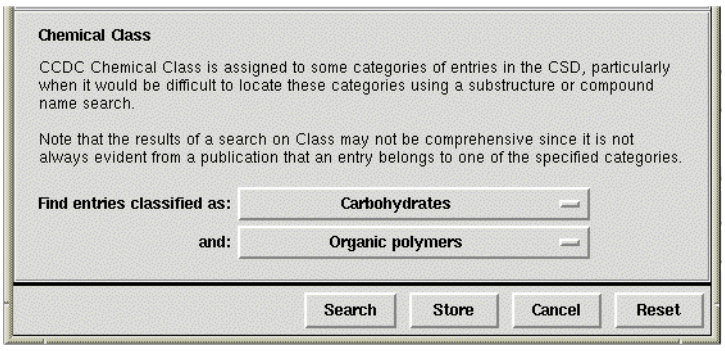


- To find structures whose names contain either *thiazol* or *fluoro*, create and store two separate queries, one for Name = *thiazol*, one for Name = *fluoro*, and then combine them (see Complex Query Combinations, page 21).
- If two separate words are typed into the same box, e.g. *benzoic acid*, a hit will only occur if the whole text string, including the space, is matched.
- Compound name searches can be combined with a chemical class search (see Combining Compound Name and Chemical Class Queries, page 118).

8.2.2 Using the Chemical Class Dialogue Box

CCDC Chemical Class is assigned to some categories of entries in the CSD to further facilitate searching. Please note that the results of a chemical class search are not comprehensive, since it is not always obvious from a publication if an entry belongs to a particular category.

- Open the dialogue box by hitting **Name/Class** in *Build Queries*.
- Select the appropriate class from the *Find entries classified as:* pull down menu, e.g. **Carbohydrates**.
- Clicking **Search** at this point will return all CSD entries classed as *Carbohydrates*.
- Alternatively a second category can be specified using the second pull-down menu e.g. **Organic Polymers**. This will return hits classed as *Carbohydrates* **AND** *Organic Polymers*.


Chemical Class
CCDC Chemical Class is assigned to some categories of entries in the CSD, particularly when it would be difficult to locate these categories using a substructure or compound name search.
Note that the results of a search on Class may not be comprehensive since it is not always evident from a publication that an entry belongs to one of the specified categories.
Find entries classified as:
Carbohydrates
and:
Organic polymers
Search
Store
Cancel
Reset

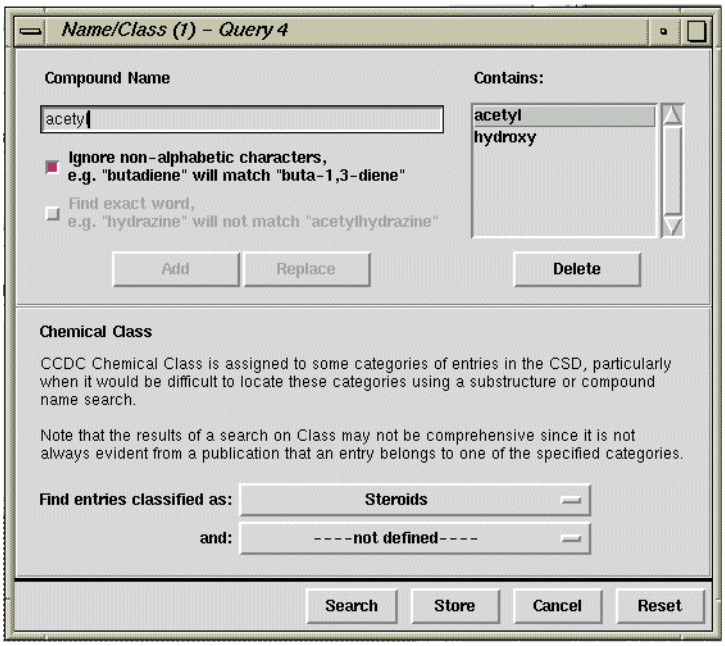
- To find structures classed as either *Carbohydrates* or *Organic Polymers*, create and store two

separate queries, one for Class=*Carbohydrate*, one for Class= *Organic Polymers*, and then combine them (see Complex Query Combinations, page 21).

- Chemical class searches can be combined with a compound name search (see Combining Compound Name and Chemical Class Queries, page 118).

8.2.3 Combining Compound Name and Chemical Class Queries

- Compound name queries can be combined with chemical class searches. The *Name/Class* dialogue box would resemble the following if searching for all CSD entries classed as *Steroids*, containing the keywords *acetyl* and *hydroxy*:


Name/Class (1) – Query 4
Compound Name
Contains:
acetyl
acetyl
hydroxy
Ignore non-alphabetic characters,
e.g. "butadiene" will match "buta-1,3-diene"
Find exact word,
e.g. "hydrazine" will not match "acetylhydrazine"
Add
Replace
Delete
Chemical Class
CCDC Chemical Class is assigned to some categories of entries in the CSD, particularly when it would be difficult to locate these categories using a substructure or compound name search.
Note that the results of a search on Class may not be comprehensive since it is not always evident from a publication that an entry belongs to one of the specified categories.
Find entries classified as:
Steroids
and:
----not defined----
Search
Store
Cancel
Reset

- To find structures classed as *Steroids*, or containing the keywords *acetyl* or *hydroxy*, create and store three separate queries, one for Class=*Steroids*, one for Name=*acetyl*, one for Name=*hydroxy*, and then combine them (see Complex Query Combinations, page 21).

8.3 Hints for Compound Name and Chemical Class Searching

- Chemical nomenclature is complex. Always consider a substructure search instead.
- You should familiarise yourself with the CSD Compound Name Conventions (see page 219).
- Compound name searching can be useful as a quick way of finding examples of a particular type of structure, since it is often quicker to type a name than draw a substructure. However, substructure searching is usually better if you want to be sure of finding all examples of a particular type of compound, since compounds may be named in unexpected ways.
- Use several short syllables rather than one long text string, e.g. *buta* and *diene* is better than *butadiene*, since the CSD will contain *1,4-butadiene* and *buta-1,4-diene*. For the same reason,

we recommend ticking the **Ignore non-alphabetic character** box.

- In general, and particularly in locating natural products, search for only the key *root* part of the name, e.g. *picolin, penicill*. This is because names may have derivative endings.
- In searching for the presence of specific elements, use only the root name of the element, e.g. *platin*. Also, remember variants such as *iron, ferr(i), ferr(o)*, etc. It is normally better to use an element search (see Searching for Specific Elements, page 120).
- Searches for trivial names, drug names, etc., can be useful. The trivial name is usually the only name stored for natural products such as alkaloids and terpenes.
- Names of certain compounds will always contain a keyword (see Compound Name Keywords Indicating Structure Type, page 220), included as a search aid, e.g. *catenane, fullerene*.
- Names of hydrates will contain the words *monohydrate, dihydrate*, etc. If other solvents are present, the name will contain the word *solvate*.
- Deuterated species will always contain the name root *deuter.*
- It is not yet possible to search for an exact compound name, a search for an exact compound name will also retrieve all hits in which the exact compound name you specified is included, so, for example, a search for *ferrocene* will also retrieve *acetyl-ferrocene*. The **Find exact word only** check box within the *Name/Class* dialogue does provide the option to search on an exact word, e.g. a search on *hydrazine* will not match *acetylhydrazine*, however, this does not represent an exact compound name search.
- A chemical class search for e.g. *Steroids* is not a comprehensive search for all steroids stored in the CSD. The *Class* field is added to an entry only if it is obvious from a publication that the structure belongs in a particular category. If searching for a specific entry that is not returned in a *Class* search, an alternative search method (e.g. a substructure search) may be more successful.

9 CHEMICAL ELEMENT SEARCHES

9.1 Typical Uses of Chemical Element Searching

- Finding structures containing particular elements.
- Finding structures containing one of a group of elements, e.g. any transition metal.
- Limiting the element types that may be present.

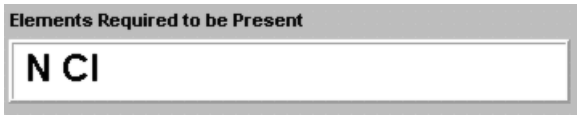
9.2 Using the Elements Dialogue Box

Open the dialogue box by hitting **Elements** in *Build Queries***.** You can:

- Search for structures containing specific elements (see Searching for Specific Elements, page 120).
- Search for structures containing one of a group of elements, e.g. any halogen (see Searching for Any of a Group of Elements, page 121).
- Specify the heaviest element that may be present in hit structures (see Specifying the Heaviest Allowed Element, page 121).

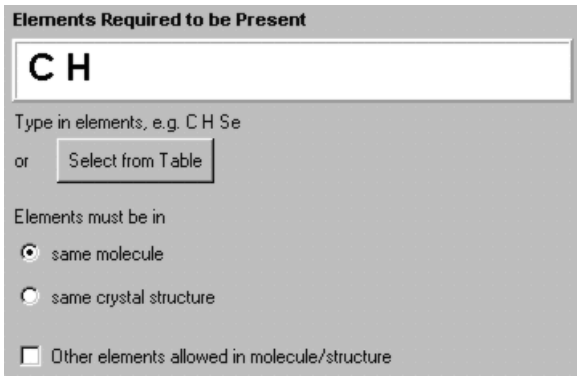
9.2.1 Searching for Specific Elements

- Open the *Elements* dialogue box.
- Type the chemical symbols of the required element(s) into the input box. Symbols must be separated by at least one space.
- Alternatively, hit **Select from Table** and pick the required elements from the Periodic Table.
- If two or more elements are specified, they must *all* be present. The example below would find molecules or structures containing *both* nitrogen and chlorine:



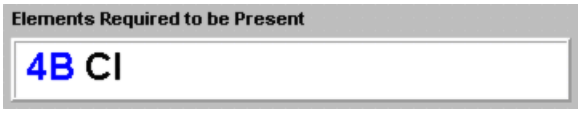
- Use the buttons to specify whether the chosen elements must be present in the **same molecule** (or ion) or in the **same crystal structure**. For example, acetylcholine perchlorate contains two ions, acetylcholine = (C_7 H_{16} N O_2) and perchlorate = (ClO_4). It would be found by the example search above if the **same structure** option were chosen, but not if the **same molecule** option were chosen (since the structure contains both N and Cl, but not in the same ion).

- If the check button **Other elements allowed in molecule/structure** is turned off, then no element other than those specified may be present. The example below would find benzene *($C_6 H_6$)* but not aniline *($C_6 H_7 N$)* or glucose *($C_6 H_{12} O_6$)*:



9.2.2 Searching for Any of a Group of Elements

- To find structures containing, e.g., either *N* or *Cl*, create and store two separate queries, one for Element = *N*, one for Element = *Cl,* and then combine them (see Complex Query Combinations, page 21).
- Alternatively, some pre-defined groups of elements may be specified by typing Element Group Symbols (see page 225) or by hitting **Select from Table** and picking the orange buttons.
- The example below would find molecules or structures containing *Cl* and any of the Group 4B elements *Ti, Zr* or *Hf:*

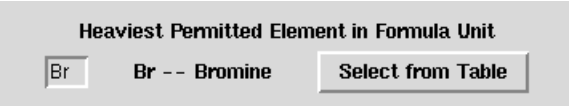


- The element groups are pre-defined in the CSD and cannot be changed, but there is a way to define and search for any required combination of elements in the *Draw* window (see Setting Variable Element Types, page 56).

9.2.3 Specifying the Heaviest Allowed Element

- Specifying the heaviest allowed element is a way of avoiding imprecise structures. This is because there may be large experimental errors in light-atom positions when a structure containing a very heavy element is determined by X-ray diffraction.
- To specify the heaviest allowed element, type the required element symbol into the appropriate input box in the *Elements* dialogue box. Alternatively, you can avoid typing by hitting **Select from Table** and picking from the Periodic Table.
- The search will reject any structure containing an element heavier than the one specified. The example below would reject any structures containing an element heavier than *Br.* This does not

mean that Br *must* be present, only that nothing heavier than Br can be present.



9.3 Hints for Chemical Element Searching

- Normal (Periodic Table) element symbols and CSD preset element group symbols may be used (see Element Group Symbols, page 225).
- Use the **same molecule** button to ensure that all specified elements are in the same bonded moiety.
- Switching off the button **Other elements allowed in molecule/structure** restricts hits to contain only the specified elements.
- The **Heaviest permitted element in Formula Unit** option can be used to eliminate structures containing heavy atoms. This is useful because the presence of heavy atoms can reduce the precision with which light atoms are located in an X-ray structure.

10 CHEMICAL FORMULA SEARCHES

10.1 Typical Uses of Chemical Formula Searching

- Finding structures with particular molecular formulae, e.g. C6 H12 O6.
- Limiting the number of atoms of specific type(s) that may be present, e.g. at least two osmium atoms.

10.2 Using the Formula Dialogue Box

Open the dialogue box by hitting **Formula** in *Build Queries*. You can:

- Specify the chemical formula to be searched for (see Specifying a Chemical Formula, page 123).
- Specify whether the formula is that of a molecule or a complete crystal structure (see Specifying Whether the Formula Applies to a Molecule or a Crystal Structure, page 124).
- Specify whether any other elements are allowed, apart from those contained in the formula you have specified (see Specifying Whether Other Elements are Allowed, page 125).
- Special considerations apply if it is desired to retrieve polymeric structures (see Hints for Chemical Formula Searching, page 125)

10.2.1 Specifying a Chemical Formula

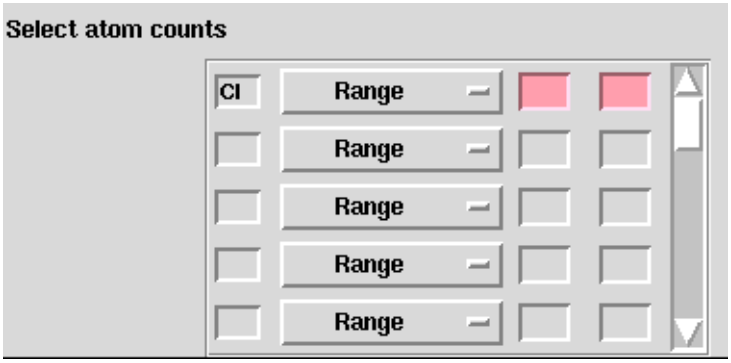
Open the *Formula* dialogue box. There are two ways of entering the required formula:

Method 1

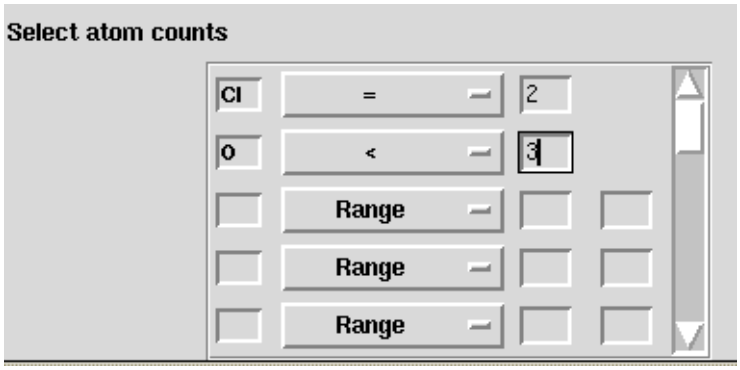
- Type directly into the input box.
- The formula may be exact, e.g. *C12 O4 H18*, or inexact, e.g. *C6-8 O > 2* (which means 6, 7 or 8 carbon atoms, and more than 2 oxygen atoms).
- Element Group Symbols (see page 225) may be used, but must be enclosed in square brackets, e.g. *C6 [7A]2* (7A means any halogen atom, so this would find molecules or structures containing 6 carbons and any two halogen atoms).

Method 2

- Hit **Select from Table** and set the formula up from the Periodic Table.
- Click on an element or any of the Element Group Symbols (see page 225). The chosen element/ element-group symbol appears in the list headed *Select atom counts*:



- Then select the required mathematical operator (see Mathematical Operators, page 18) and type in the number(s) of atoms that must be present.
- Continue until all elements and atom counts are defined. The example below specifies a formula containing 2 chlorine atoms and less than 3 oxygens:



- Hit **Done** when the required formula is specified.

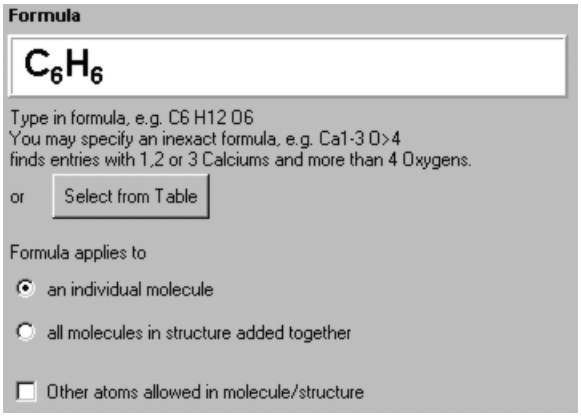
10.2.2 Specifying Whether the Formula Applies to a Molecule or a Crystal Structure

- In searching for a chemical formula, you can specify whether the formula is that of a single molecule, or of all the molecules in the structure added together. For example, the crystal structure acetylcholine perchlorate contains two individual ions with formulae C7 H16 N1 O2 (acetylcholine) and Cl1 O4 (perchlorate); the formula of the individual ions added together is C7 H16 N1 O6 Cl1.
- Choose one of the buttons **an individual molecule** or **all molecules in structure added together** to select whether the specified formula applies to an individual molecule or ion, or all the molecules or ions in the structure added together.

10.2.3 Specifying Whether Other Elements are Allowed

In searching for a chemical formula, you can specify whether any other elements are allowed, apart from those contained in the formula you have specified:

- If the check button **Other atoms allowed in molecule/structure** is turned off, then no element other than those in the specified formula may be present. The example below would find benzene, *C6 H6,* but not phenol *C6 H6 O* because the **Other atoms allowed in molecule/structure** option is switched off:


Formula
C6H6
Type in formula, e.g. C6 H12 O6
You may specify an inexact formula, e.g. Ca1-3 O>4
finds entries with 1,2 or 3 Calciums and more than 4 Oxygens.
or
Select from Table
Formula applies to
an individual molecule
all molecules in structure added together
Other atoms allowed in molecule/structure

10.3 Hints for Chemical Formula Searching

- A crystal structure may consist of more than one molecule. In such circumstances each molecule is formulated separately, e.g. *C6 H5 N2 O3, H2 O1* (see Chemical Formula, page 218). To do an exact search for this structure, construct separate queries for each of the two molecular formulae and combine them, i.e. search for entries that have both (see Simple Query Combinations, page 20).
- If polymeric structures are to be retrieved then it is important to apply the formula to **all structures in the molecule together** and not to an **individual molecule**. The reason is that the latter search uses the information in the 2D diagram. There are cases where atoms from adjacent monomeric units are included in the 2D diagram to illustrate connectivity (see for instance the extra Mo atoms in AXABEG and BALFAV). The former search will only look at the constitution of single monomeric units.

11 SPACE GROUP AND CRYSTAL SYSTEM SEARCHES

11.1 Typical Uses of Space Group and Crystal System Searching

- Finding all entries in a particular crystal system, e.g. tetragonal.
- Finding all entries in a particular space group, e.g. P21/c.
- Restricting other searches to specific crystal systems or space groups.

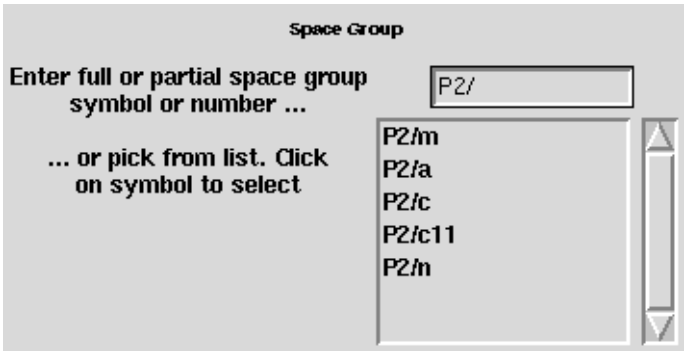
11.2 Using the Space Group Dialogue Box

Open the dialogue box by hitting **Space Group** in *Build Queries*. You can:

- Search for a particular space group (see Searching for a Space Group, page 126).
- Specify whether alternative space-group settings are allowed (see Including or Excluding Alternative Space-Group Settings, page 127).
- Search for a range of space groups (see Searching for a Range of Space Groups, page 127).
- Limit the search to Sohncke (chiral) space groups (see Restricting the Search to Sohncke or Chiral Space Groups, page 128).
- Search for a particular crystal system (see Searching for Crystal Systems, page 128).

11.2.1 Searching for a Space Group

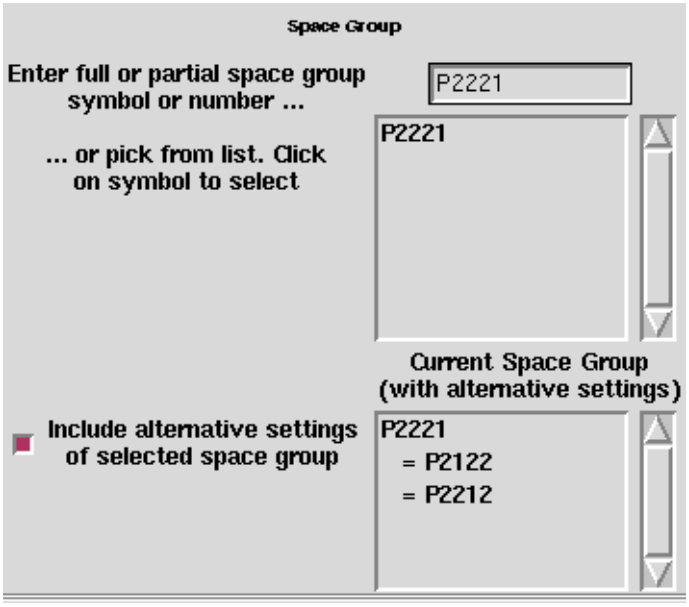
- Open the *Space Group* dialogue box.
- Type the space group number (valid range 1-230) or the space group symbol (see Space Group, page 240) into the input box.
- Alternatively, pick the space group symbol from the top list box, either by scrolling down the list with the scroll bar, or by typing the first character or two of the space group symbol into the input box. As the symbol is typed, the list in the box is narrowed down:



- Pick the required symbol with a single mouse click.

11.2.2 Including or Excluding Alternative Space-Group Settings

- If the check-box **Include alternative settings of selected space group** is switched on, the search will find all occurrences of the space group that has been defined, even those with a different setting (i.e. different choice of axes) (see Space Group, page 240).
- The different settings of the space group will be displayed in the lower pick list; e.g. if *P2221* has been typed in the input box, the lower list will contain the alternative settings *P2122* and *P2212*:


Space Group
Enter full or partial space group symbol or number ...
P2221
... or pick from list. Click on symbol to select
P2221
Current Space Group (with alternative settings)
Include alternative settings of selected space group
P2221
= P2122
= P2212

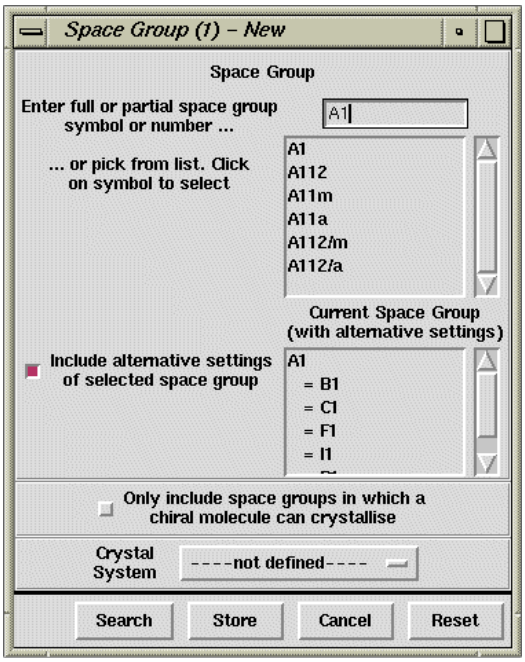
- Since the choice of space-group setting is arbitrary, it will be usual to leave this check-box switched on.

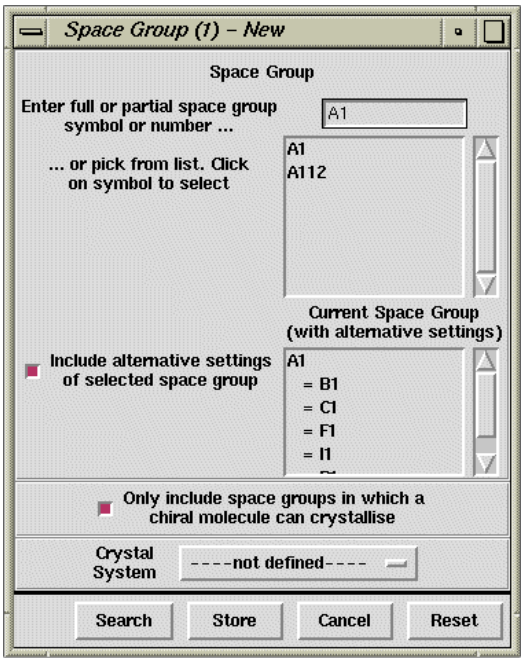
11.2.3 Searching for a Range of Space Groups

- If only part of a space group symbol is typed in, the search will find all space groups beginning with the specified string of characters, e.g. *P21/* would find all of *P21/m, P21/a, P21/c*, etc.
- However, *P21*, being a valid symbol in its own right, would find *only* this particular space group.
- To find structures in either of, e.g., *P21* and *P21/m*, create and store two separate queries, one for Space Group = *P21* and one for Space Group = *P21/m*, and combine them (see Complex Query Combinations, page 21).
- Searches can be made on aspect symbols (symbols for incompletely determined space groups, where the character * is used as a wild-card character, e.g. P*aa) (see Space Group, page 240).

11.2.4 Restricting the Search to Sohncke or Chiral Space Groups

- This option can be switched on by activating the *Only include space groups in which a chiral molecule can crystallise* tick box (see Sohncke Groups (Chiral Space Groups), page 240). The space group window will update depending on whether the tick box is on or off.





11.2.5 Searching for Crystal Systems

- Open the *Space Group* dialogue box.
- All structures belonging to a particular crystal system can be found by clicking on the button labelled **--- not defined ---** and selecting the required crystal system from the resulting pull-down menu.
- It is unnecessary to define a crystal system if a space group has already been specified.

11.3 Hints for Space Group Searching

- It is almost never sensible to switch off the check-box labelled **Include alternative settings of selected space group**. In many space groups, it is possible to choose the unit cell axes or origin in more than one way (e.g. *Pmm2, P2mm and Pm2m* are different settings of the same space group) and the choice between them is essentially arbitrary (see Space Group, page 240).
- Rhombohedral space groups can be expressed relative to hexagonal or rhombohedral axes. If rhombohedral axes have been used, the letter *r* is appended to the space group symbol, e.g. *R3r*.

12 UNIT CELL SEARCHES

12.1 Typical Uses of Unit Cell Searching

- Finding structures with specific Reduced Cell Parameters (see page 238) to find if a new crystal has already been studied.
- Locating all structures with short axial lengths, perhaps in a crystal engineering context.

12.2 Using the Unit Cell Dialogue Box

Open the dialogue box by hitting **Unit Cell** in *Build Queries*.

- On entering the dialogue box, it is first necessary to specify whether a reduced cell search is required (see Reduced Cell Parameters, page 238).
- The answer will be *yes* if the purpose of the search is to find out whether a crystal structure with known cell dimensions is present in the CSD, e.g. when checking whether a particular structure has already been determined.
- The answer will be *no* in a few unusual situations, e.g. to find structures with a particular repeat distance along an axis.

12.2.1 Reduced Cell Searching

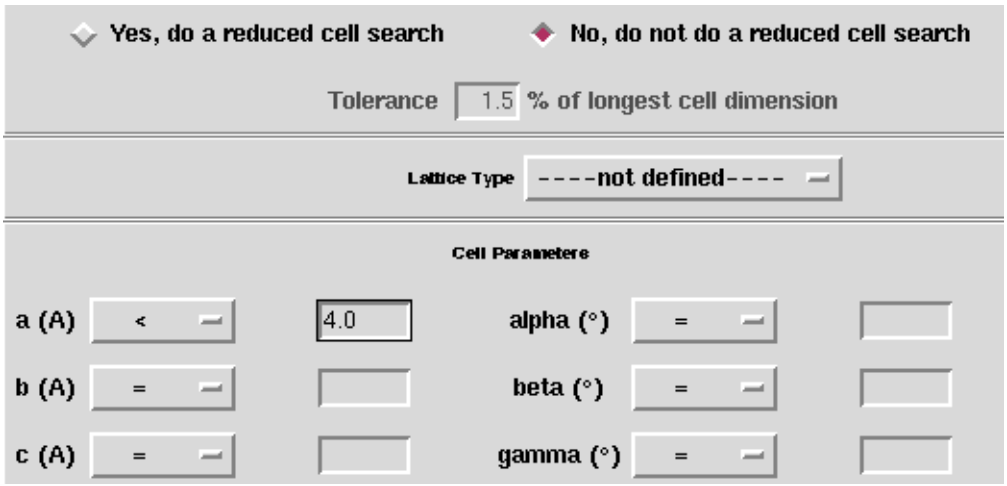
Choose by clicking on **Yes: do a reduced cell search:**

- It is essential to specify the Lattice Type (see page 231) and all the cell parameters of the cell to be searched for.
- Select the *Lattice Type* from the pull-down menu obtained by hitting the button labelled **--- not defined ---**.
- Type the cell parameters (a, b, c, alpha, beta, gamma) into the input boxes (see Unit Cell Parameters, page 242).
- A tolerance value is required. This is expressed as a percentage of the longest cell dimension and determines how much bigger or smaller a cell parameter in the CSD may be, compared with the input value, and still be considered a match.
- The tolerance should reflect the accuracy of the input unit cell data, a larger value being necessary if the cell parameters are imprecise.
- The default tolerance value (1.5%) has been selected to allow for typical experimental errors, and for the average difference in cell dimensions observed between determinations at room and reduced temperatures.

12.2.2 Searching for Non-Reduced Cell Parameters

Choose by clicking on **No: do not do a reduced cell search:**

- The lattice type may be set or left undefined.
- Any or all of the cell parameters a, b, c, alpha, beta, gamma may be typed into the input boxes (see Unit Cell Parameters, page 242), with appropriate mathematical operators (see Mathematical Operators, page 18) chosen from the pull-down menus.
- The example below would find structures with *a* cell lengths less than *4.0* Å:



- Because of the possibility of different space group settings, it may be necessary to set up, store and combine more than one query. For example, finding all crystal structures with a repeat distance below 4.0A requires three queries, one for a < 4.0, one for b < 4.0 and one for c < 4.0. These must be combined (see Complex Query Combinations, page 21) to find all structures that satisfy at least one of the criteria.

12.3 Hints for Unit Cell Searching

- Since the choice of cell axes is often arbitrary, it is normally better to search on the Reduced Cell Parameters (see page 238).
- Reduced cell parameters are subject to experimental uncertainties. Thus, all searches should specify a reasonable tolerance value. The default tolerance has been chosen carefully and will normally be suitable.
- It is essential to specify correctly the Lattice Type (see page 231) of your crystal when doing reduced-cell parameter searching.
- When viewing the results of a reduced-cell search, do not be surprised if the reduced-cell dimensions of hit structures are different from the cell parameters you typed in. This will be because your parameters have themselves been reduced to standard form.
- Reduced cell searching only involves comparison of cell lengths, not angles. This is because the angles can be numerically unstable.

13 CRYSTAL DENSITY AND UNIT-CELL CONTENT SEARCHES

13.1 Typical Uses of Density and Unit-Cell Content Searching

- Finding crystals with densities in a particular range.
- Finding crystals with unusually high or low densities.
- Finding crystals with specific/unusual Z values.
- Finding crystals with specific/unusual Z' values.
- Finding crystals with two or more crystallographically-independent molecules (i.e. Z' > 1).
- Finding molecules that lie on crystallographic symmetry elements (i.e. Z' < 1).
- Finding crystals containing a specific number of chemical units (molecules, ions, etc.).
- Limiting the number of atoms in a structure (e.g. to exclude very small structures or very large structures).

13.2 Using the Z/Density Dialogue Box

Open the dialogue box by hitting **Z/Density** in *Build Queries*.

- *Z (Number of Molecules per Unit Cell)* (see page 246)**:** Select the required mathematical operator (see Mathematical Operators, page 18) and type a number into the input box. In this context, *Molecule* means the whole formula unit, which may actually consist of more than one chemically discrete entity, e.g. an ion pair or a solvated structure.
- *Z' (Number of Molecules per Asymmetric Unit)* (see page 247): Select the required mathematical operator (see Mathematical Operators, page 18) and type a number into the input box. In this context, *Molecule* means the whole formula unit, which may actually consist of more than one chemically discrete entity, e.g. an ion pair or a solvated structure.
- *Number of Atoms with 3D Coordinates* (see page 234): Select the required mathematical operator (see Mathematical Operators, page 18) and type a number into the input box.
- *Calculated Density* (see Density, page 223): Select the required mathematical operator (see Mathematical Operators, page 18) and type a number into the input box. The example below would find all structures with density > *2.0* g/cc:

- *Number of Chemical Units (molecules, ions, etc.)* (see page 234): Select the required mathematical operator (see Mathematical Operators, page 18) and type a number into the input box.

13.3 Hints for Density and Unit-Cell Content Searching

- Measured densities are sometimes reported in papers but tend to be unreliable and are no longer stored in the CSD.
- Any search for a density should allow a reasonable tolerance, i.e. use *range* rather than =.
- Number of Atoms with 3D Coordinates (see page 234): This number might not match exactly the number of atoms in the chemical formula (if, for example, not all of the H atoms were located experimentally).
- Structures with Z'>1 are interesting because they contain two or more crystallographically independent molecules, i.e. molecules which have the same chemical structure but are not forced by symmetry to have identical geometries. Comparing such molecules often gives clues about molecular flexibility.
- Any search in which the number of chemical units or residues is specified is best used in conjunction with other tests to exclude, for example, multi-residue entries containing solvates or ionic compounds.

14 SEARCHES FOR EXPERIMENTAL DETAILS

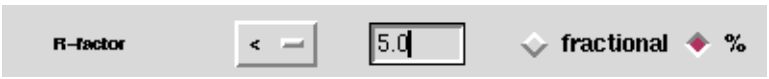
14.1 Typical Uses of Searching for Experimental Details

- To find structure determinations carried out at high or low temperatures.
- To find structures with accurate hydrogen-atom positions (i.e. determined by neutron diffraction).
- To ensure that only precise structures are found, or to exclude disordered structures or those with unresolved errors.

14.2 Using the Experimental Dialogue Box

Open the dialogue box by hitting **Experimental** in *Build Queries***.**

- *R-factor* (see page 239): Choose whether to input as a decimal number (valid range 0-1) or a percentage (valid range 0-100). Select the required mathematical operator (see Mathematical Operators, page 18); the usual choice will be **<** or **<=**, since structures with low R factors are the most precise. Type a value into the input box:



- **Exclude disordered structures** (see Disordered Structures, page 224): Turn on the check button to exclude structures with crystallographic disorder. This can also be done in the *Filters* section (see FILTERS: SECONDARY SEARCH CRITERIA, page 143) of the *Search Setup* box.
- **Exclude structures with unresolved errors** (see Unresolved Errors, page 243): Turn on the check button to exclude structures containing unfixed errors. This can also be done in the *Filters* section (see FILTERS: SECONDARY SEARCH CRITERIA, page 143) of the *Search Setup* box.
- *Average e.s.d. of C-C Bonds* (see page 214): Select **Any** (allows all structures) or **< 0.005**Å (allows only precisely determined structures).
- **Exclude powder structures** (see Powder Data, page 237): Turn on the check button to exclude structures determined from powder diffraction studies. This can also be done in the *Filters* section (see FILTERS: SECONDARY SEARCH CRITERIA, page 143) of the *Search Setup* box.
- *Temperature of Structure Determination* (see page 241)**:** Choose units (degrees K or C). Select the required mathematical operator (see Mathematical Operators, page 18). Type a value into the input box:

Alternatively, select **=room** to find structures determined at room temperature (283-303K).

- *Radiation source*: Select **Any** to allow all structures, **X-Rays** to restrict to structures determined by X-ray diffraction, **Neutron** to restrict to structures determined by Neutron Diffraction (see page 233).

14.3 Hints when Searching for Experimental Details

- R-factor: In general, the lower the R-factor, the better the structure. Look at the Glossary entry on R-factor (see page 239) for a rough guide to the relationship between R-factor and experimental quality.
- Disordered Structures (see page 224): When coordinates for two alternative sites are reported, those corresponding to the site of minor occupancy are *suppressed* in the CSD. If a set of atoms is disordered equally over two sites, one set is arbitrarily retained and the other is suppressed.
- Disordered parts of structures tend to be very imprecise and should not be used for geometrical analysis. In many cases, however, only the solvent molecules in a structure are disordered and the molecule of major relevance is well characterized.
- Unresolved Errors (see page 243): If a search yields very few hits it may be worth relaxing the conditions to allow entries with unresolved errors. The errors may relate to irrelevant parts of the structures such as solvent molecules.
- Average e.s.d. of C-C Bonds (see page 214): Structures with average e.s.d < 0.005Å are of high experimental quality.
- Structures determined by neutron diffraction are useful because the hydrogen atom positions are much more accurate than those in X-ray structures.

15 GENERAL TEXT SEARCHES

15.1 Typical Uses of Text Searching

- Find crystals of a specific colour.
- Find structures whose absolute configuration was determined by X-ray methods.
- Look for the words *drug*, *activity*, *agent* etc. to find biologically, pharmaceutically or agrochemically active molecules.
- Search for particular experimental conditions such as *neutron, synchrotron, powder.*
- Find polymorphs by searching for the word *polymorph.*
- Find crystals with specific morphology, e.g. *rhombs, acicular,* etc.

15.2 Using the All Text Dialogue Box

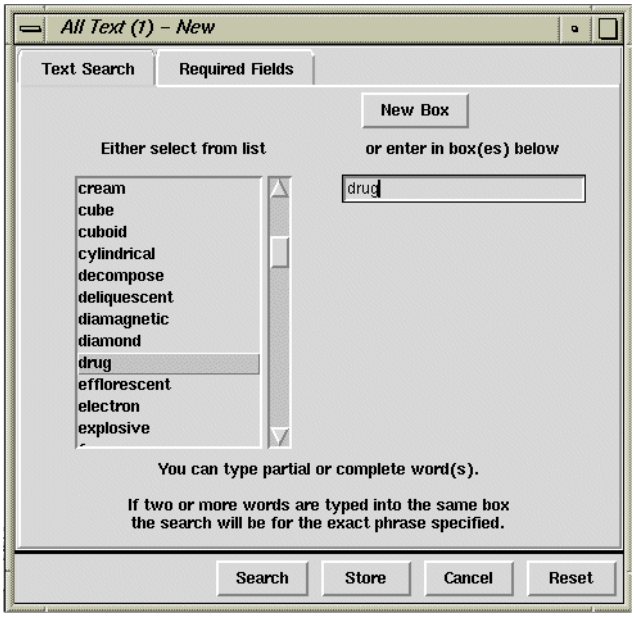
The All Text dialogue box has two tabs:

- *All Text* (see All Text Searching, page 135)
- *Required Fields* (see Specifying Required Fields, page 137)

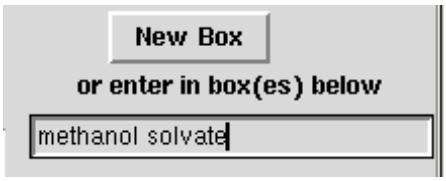
It is also possible to combine *All Text* and *Required Field* searches (see Combining All Text and Required Fields Searches, page 138).

15.2.1 All Text Searching

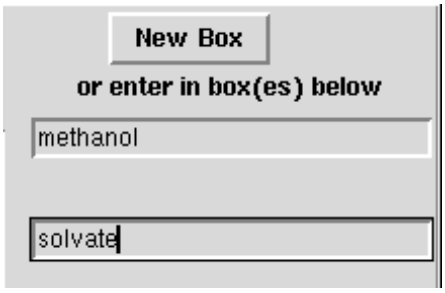
- Open the dialogue box by hitting **All Text** in *Build Queries*.
- Either select the desired keyword from the scrolling list (please note that this is not an exhaustive list) or type directly into the input box. The example below will find all entries containing the keyword *drug*:



- If two or more words are typed into the same input box, the search will find only those entries containing the exact phrase specified. The example below will find entries containing the exact sequence of characters *methanol solvate* (i.e. these two words separated by a single space):



- To find entries containing two or more words that need not be adjacent, use the **New Box** button and type the required words into separate input boxes. The example below will find entries containing the words *methanol* and *solvate*, but these words do not need to be next to each other in the text comment:



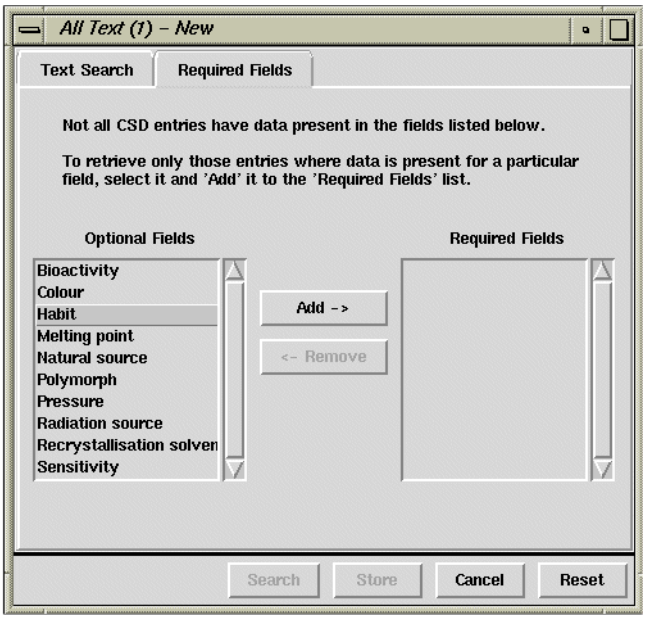
- To find entries containing either of two keywords (e.g. *needle* or *plate*), create and store two separate queries and combine them (see Complex Query Combinations, page 21).

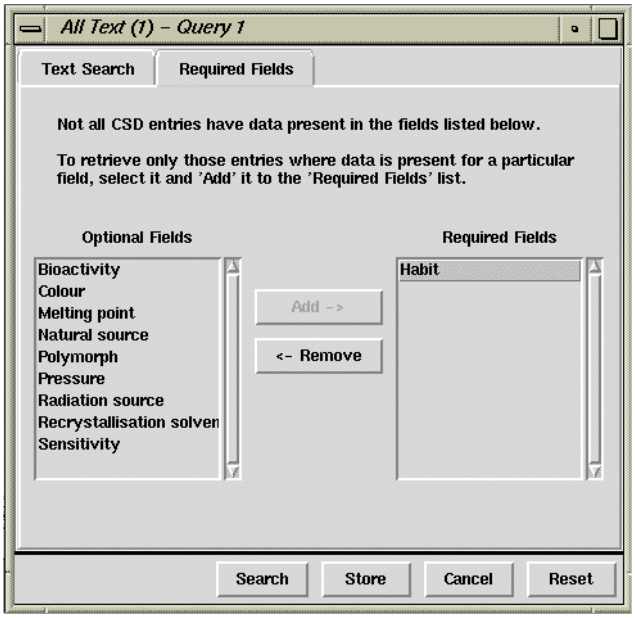
- *All Text* and *Required Field* searches can be combined in an AND search within the *All Text* dialogue box (see Combining All Text and Required Fields Searches, page 138).

15.2.2 Specifying Required Fields

This sort of search will retrieve hits based on information being present in particular fields, e.g. crystal colour, habit etc. Field information is stored with the CSD entry if it has been provided in the publication, thus some entries may not have details present in particular fields. It is possible to combine a *Required Fields* search with an *All Text* search to refine the hitlist obtained (see Combining All Text and Required Fields Searches, page 138).

- Open the dialogue box by hitting **All Text** in *Build Queries*, then click on the **Required Fields** tab.
- Select the desired keyword from the *Optional Fields* scrolling list and click **Add** to create an entry in the *Required Fields* box. The example below will find all entries containing information in the *Habit* field:



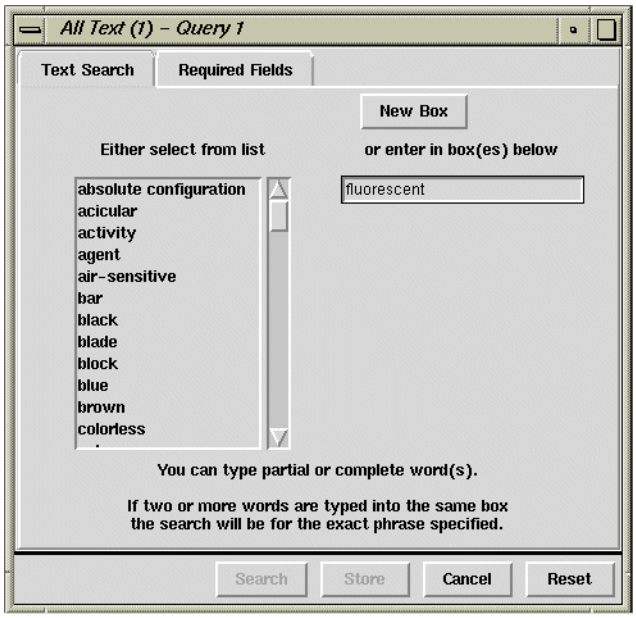


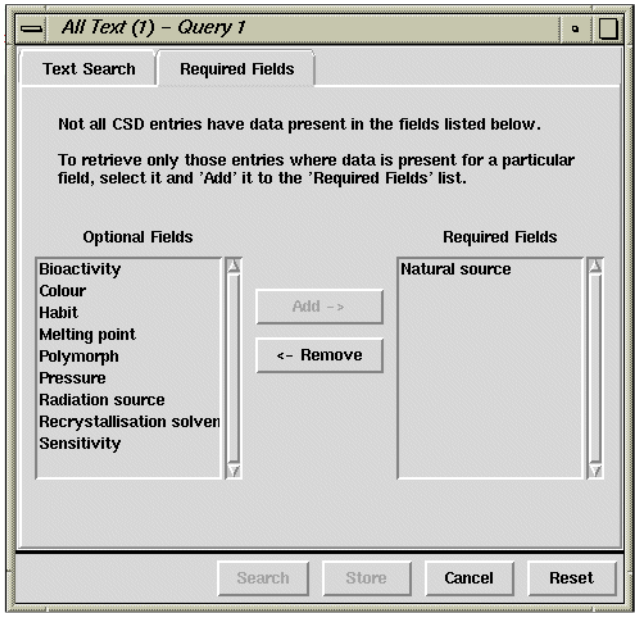
- To find all entries that contain *Habit* and crystal *Colour* information, select **Colour** then **Add**. This will result in *Habit* and *Colour* entries in the *Required Fields* list.
- To find entries that contain either *Habit* or crystal *Colour* information, create and store two separate queries and combine them (see Complex Query Combinations, page 21).
- To remove a field from the *Required Fields* list, select the appropriate word then click on **Remove**.

15.2.3 Combining All Text and Required Fields Searches

All Text and *Required Fields* searches can be combined in an AND search within the *All Text* dialogue box.

- The following search will return all entries containing the keyword *fluorescent* **and** information present in the *Natural Source* field:





- To find entries containing either a keyword (e.g. *fluorescent*) or keywords, **or** the presence of information content in a CSD field (e.g. *Natural Source* information), create and store separate queries for each keyword/field and combine them (see Complex Query Combinations, page 21).

15.3 Hints for All Text/Required Field Searching

- Use of the **All Text** tab will cause *all* of the textual information to be searched. This includes some information which can be searched more quickly and effectively through other parts of the interface, e.g. author names, journal information, compound names and formulae.
- Use the All Text option combined with the Required Field words to find out about absolute configuration, colour, crystal form (i.e. polymorphism), drug activity, crystal habit, miscellaneous physical and chemical properties, how crystals were grown, etc.
- Browse through the scrolling lists in both dialogue boxes to get ideas for useful search terms or available CSD fields. Note that the scrolling list is not exhaustive: if a chosen keyword isn't present, it can still be searched for by typing it in the accompanying dialogue box.
- The All Text option *can* be used for finding entries by specifying the Deposition Numbers (see page 224) that are printed in some journals. For example, *CCDC 182/9* is a deposition number indicating a structure published in *Chem.Comm.*; this number is cited in the paper.

Comprehensive deposition number searching is provided through the Author/Journal dialogue box, accessed through the Build Queries window (see CCDC number (deposition number), page 115).

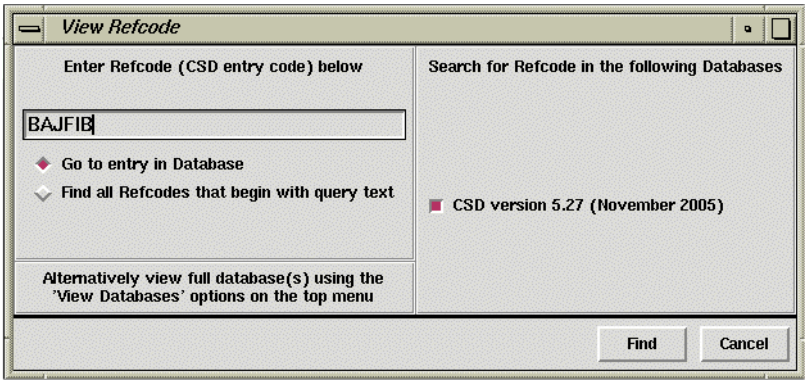
- A text search for *form* will produce some unexpected hits, e.g. electron de*form*ation density, multiple de*form*ation, charge de*form*ation, con*form*ation, con*form*er.
- Searches for *activity* or *agent* will also produce some unwanted hits if the *Required Field* keyword *Bioactivity* is not defined, e.g. photochemical *activity*, dye *agent*.
- English spelling is used in most cases, e.g. *hypoglycaemic, tumour* however American spelling is kept if it was used in the original publication.
- Parts of words or the root of the word can be used for searching, e.g. if searching for the keyword *immunosuppressant* or *immunosuppressive* compounds, use *immunosuppress*.
- Performing a *Required Fields* search will only return entries where there is information present in the chosen field. Not all CSD entries contain information in the listed fields. This information is generally gleaned from the publication citing the structure: we do not actively seek further information ourselves.

16 REFCODE SEARCHES

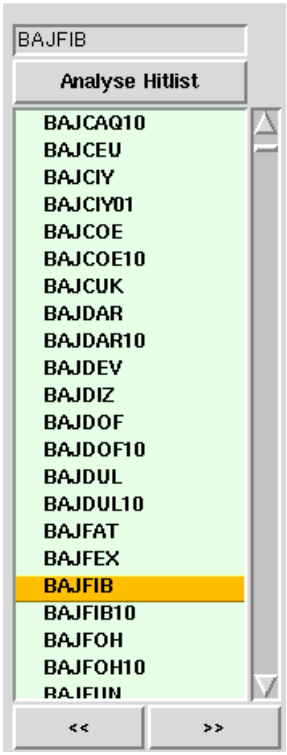
16.1Exact Refcode Searching

There are two ways of finding an entry whose Refcode (i.e. CSD entry identifier) is known (see Refcode (entry ID), page 238):

- Hit the **Refcode (entry ID)** button in *Build Queries* and type the required Refcode into the input box, e.g.


View Refcode
Enter Refcode (CSD entry code) below
Search for Refcode in the following Databases
BAJFIB
Go to entry in Database
Find all Refcodes that begin with query text
CSD version 5.27 (November 2005)
Alternatively view full database(s) using the 'View Databases' options on the top menu
Find
Cancel

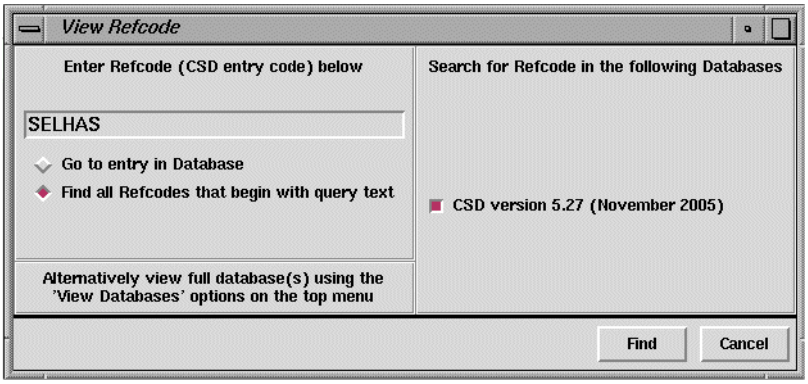
- Ensure the **Go to entry in Database** radio button is activated and hit **Find**.
- The Refcode will consist of six letters and, possibly, two numbers.
- Select **View Databases** from the top-level menu. This enables the entire CSD to be loaded. A particular entry may then be viewed by typing its Refcode into the grey box at the top of the Refcode list in *View Results* or by scrolling down the list and clicking on the desired Refcode:



16.2 Searching for Refcode Families and Multiple Structure Determinations

It is possible to search for Refcode families or structures in the CSD that have been determined multiple times.

- Hit the Refcode (entry ID) button and enter e.g. **SELHAS**.


View Refcode
Enter Refcode (CSD entry code) below
SELHAS
Go to entry in Database
Find all Refcodes that begin with query text
Alternatively view full database(s) using the 'View Databases' options on the top menu
Search for Refcode in the following Databases
CSD version 5.27 (November 2005)
Find
Cancel

- Ensure the *Find all Refcodes that being with query text* radio button is activated and click **Find**.
- This will return all CSD entries with the query Refcode i.e. a refcode family.

- The results of these Refcode searches can be combined in the *Manage Hitlists* window of ConQuest (see Managing Hitlists, page 23).
- It is also possible to search for refcodes which contain less than six characters e.g. CYCBO. This search will return all refcodes with the query text, however please note that these refcodes are not necessarily related to one another.

16.3 Saving and Reading Refcode Lists

- To save a list of Refcodes, ensure that it is currently on display in *View Results* (if it is not, display it by using the **Results** top-level menu option, then select **Searches** from the resulting pull-down menu to display the search of your choice (see Switching between Searches, page 40).
- Pick **File** from the top-level menu followed by **Export Entries as...** in the resulting pull-down menu, then click on the file type bar and select **Refcode: csd entry identifier list** from the pull-down menu to save:
  - the individual structure currently being displayed (turn on **Current entry only**), or
  - all unsuppressed structures in the hit list (turn on **All selected entries**) (see Suppressing Unwanted Entries, page 40).
- Saved Refcode lists will be given the extension *.gcd* which can be read back into ConQuest and displayed as if the search had just been run.
- To read a previously saved Refcode list, pick **File** from the top-level menu, followed by **Open** and then select **Refcode List...** from the resulting pull-down menu. Loading a refcode list for the first time may result in a progress window. If you have previously loaded Refcode lists, resulting from searches in the same ConQuest session, these can be viewed by selecting **Results** from the top-level menu followed by **Refcode Lists** in the resulting pull-down menu to display the Refcode list of your choice.
- Alternatively, you can read in a Refcode list, search or hitlist combination, saved from an earlier ConQuest session, by picking **File** from the top-level menu followed by **Open** and then select **Recent Files** from the resulting pull-down menu.

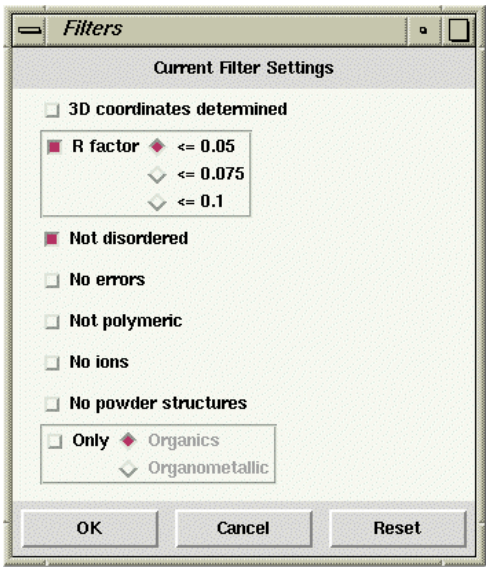
17 FILTERS: SECONDARY SEARCH CRITERIA

17.1 Typical Uses of Filters

Filters are secondary search criteria used to restrict more general searches on the basis of, for example, precision, whether disorder is present, whether there are unresolved errors in the structure, or whether the structure is organic or organometallic.

17.2 Applying Filters

- All displayed filters (see Default Filter Settings, page 144) are listed in the *Search Setup* dialogue box, which is displayed every time a search is requested.
- Click on the check button next to a filter to change its setting. In the example below, the filters **R-factor** <= 0.05 and **Not disordered** are turned on (i.e. will be used in the search about to be started), the others are turned off:



- When the *Search Setup* dialogue box is opened for the second and subsequent searches in a session, filter settings are remembered from the previous search.
- It is possible to control which filters are, by default, displayed when the *Search Setup* dialogue box is opened for the first time in a ConQuest session (see Default Filter Settings, page 144).

17.3 Available Filters

- **3D coordinates determined**: Switch on to exclude structures for which no 3D information is available.
- **R-factor** (see page 239)**:** Switch on the check button and then click on one of the three radio buttons to restrict the search to structures with R-factors less than or equal to 0.05, 0.075 or 0.1. If an R factor limit other than one of these is required, set up as a query using the *Experimental* dialogue box (see Using the Experimental Dialogue Box, page 133). If an R factor query and this filter are set at different levels, the more stringent (i.e. lower) of the two criteria will be used.
- **Not disordered:** Switch on to exclude structures which contain crystallographic disorder (see Disordered Structures, page 224). This can also be done in the *Experimental* dialogue box (see Using the Experimental Dialogue Box, page 133).
- **No errors**: Switch on to obtain only those structures which are error-free at the 0.05Å level.
- **Not polymeric**: Switch on to exclude polymers and metal catena structures (see Polymeric Bonds, page 236).
- **No ions**: Switch on to exclude structures containing ions. Zwitterions will not be excluded.
- **No powder structures**: Switch on to exclude structures in which the data was obtained from a powder diffraction study (see Powder Data, page 237). This can also be done in the *Experimental* dialogue box (see Using the Experimental Dialogue Box, page 133).
- **Only Organics** (see page 235) or **Only Organometallics** (see page 235): Switch on the check button and then click on the appropriate radio button to restrict the search to (i) organic structures only, or (ii) organometallic structures only.

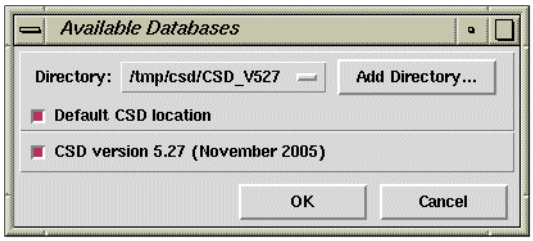
17.4 Default Filter Settings

- To control which filters (see Available Filters, page 144) are displayed in the *Search Setup* dialogue box, which is displayed every time a search is requested, select the top-level menu button **Options**, followed by **Filters...** then **Filters to Display...** in the resulting pull-down menus. To make a filter available enable the check button next to the filter. Settings are retained between ConQuest sessions.
- To control which of the visible filters are, by default, turned on when the *Search Setup* dialogue box is opened for the first time in a ConQuest session select the top-level menu button **Options**, followed by **Filters...** then **Default Filter Settings...** in the resulting pull-down menus. To apply a filter enable the check button next to the filter. Settings are retained between ConQuest sessions.

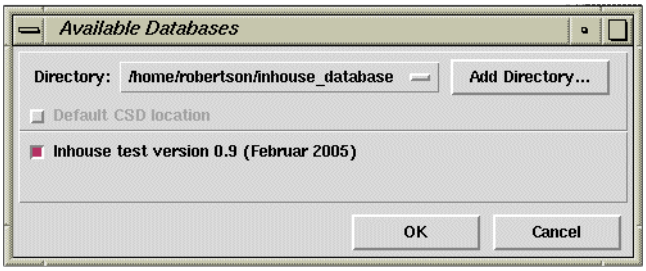
18 VIEWING AND SEARCHING IN-HOUSE DATABASES, DATABASE SUBSETS AND PREVIOUS CSD VERSIONS

18.1 Viewing In-House Databases

- If you have used the CCDC program PreQuest (see page 237) to build a database of your own crystal structures in CSD format, the database can be viewed in ConQuest.
- You must first configure ConQuest so that it can find the database (see CSDS Release & Installation Notes).
- All entries in the database can then be viewed by clicking on **View Databases** in the top level menu. This will display a list of all the databases that ConQuest has been configured to use, including the CSD itself and any in-house database(s). Select the one you want.
- If you have a configured in-house database in a directory separate to the main CSD files, the in-house database can be viewed by selecting the **Available Databases...** option from the **View Databases** menu:



- Click on **Add Directory...**, browse to the folder where the in-house database files are stored and select the **.inf* file.



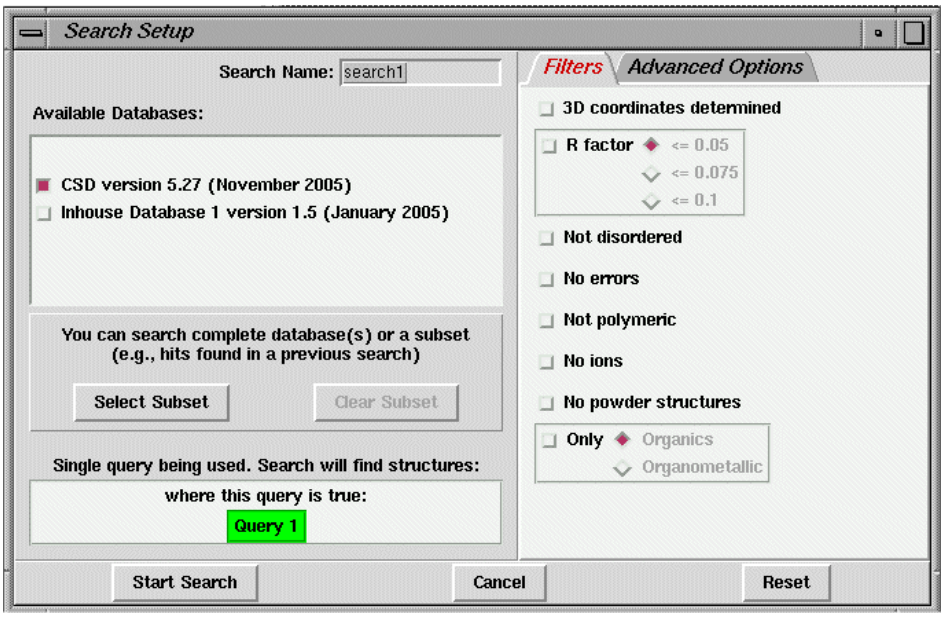
- The radio button next to the database name will need to be activated to complete the selection.
- The Directory pull-down menu can then be used to toggle between the configured databases.
- Databases can be removed from the **View Databases** pull-down menu by deselecting the radio button adjacent to their name in the *Available Databases* window.

18.2 Searching In-House Databases

- If you have used the CCDC program PreQuest (see page 237) to build a database of your own

crystal structures in CSD format, the database can be searched in ConQuest.

- You must first configure ConQuest so that it can find the database (see CSDS Release & Installation Notes).
- You can then perform searches on the in-house database, either on its own or together with the CSD.
- Construct your query as usual and hit **Search.** A list of available databases will be displayed in the resulting *Search Setup* dialogue box, e.g.



- Switch on the database(s) that you want to include in the search.
- Once you have made your database selection, hit **Start Search** to begin searching.
- ConQuest can handle more than one in-house database if necessary.

18.3 Searching Database Subsets

Rather than searching the whole of the CSD, it is possible to search a subset of the database (or a subset of both the CSD and in-house databases, should you have any of the latter). You can restrict a search to:

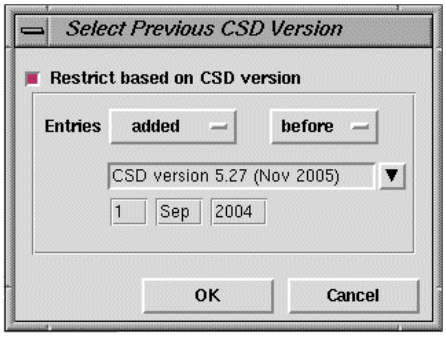
- Entries added/modified before/after a particular CSD release or a specific date (see Searching CSD Entries Added or Modified Before or After a Particular Release or Date, page 147).
- Entires in a pre-defined hitlist. This allows you to search against the only best representative of each unique polymorph in the CSD, or to restrict searches to the Classroom subset (see Searching on entires in a pre-defined hitlist, page 148).
- The hits found in a search you have already run during the current ConQuest session (see

Searching Hits from a Previous Search in the Current ConQuest Session, page 149).

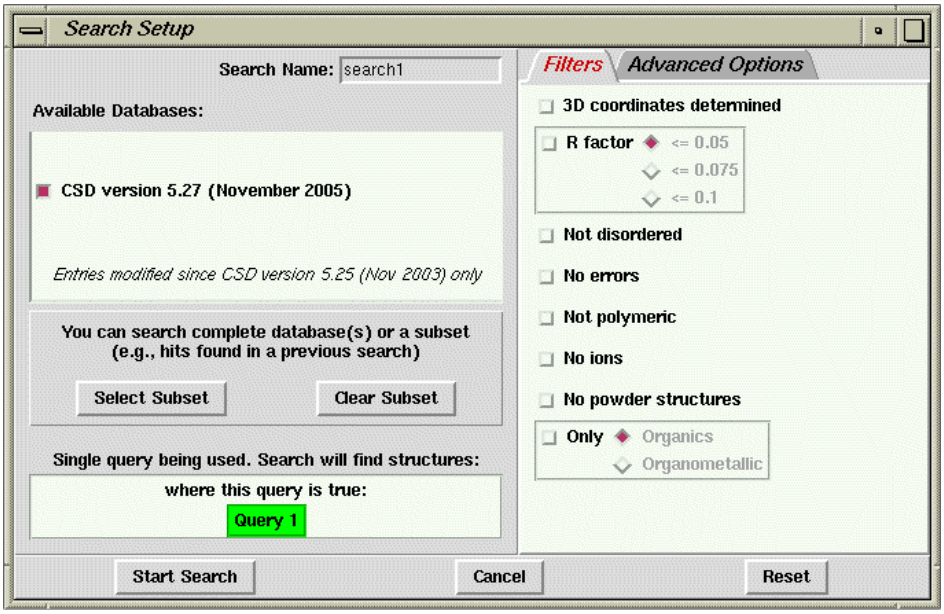
- The hits found in a search that has been saved to disk, either in native ConQuest format (*.cq*s), a hitlist combination (*.chl*), or as a refcode list (*.gcd*) (see Searching the Hits from a Search Previously Saved to Disk, page 149).

18.3.1 Searching CSD Entries Added or Modified Before or After a Particular Release or Date

- Build the query as usual and hit **Search**. Then hit the **Select Subset** button in the *Search Setup* dialogue box.
- Click on the **CSD Version...** button at the bottom left of this window. This will launch the *Select Previous CSD Version* window.
- To view the database scrolling list, click on the **Restrict based on CSD version** radio button. The window will resemble the following:



  - The *Entries* pull-down menus are used to indicate if the search should include *Entries added* **before** or **after** a particular CSD release, or if the *Entries* have been *modified* **before** or **after** a particular CSD release.
  - Select the CSD version of interest from the central pull-down menu, e.g. *CSD version 5.25 (Nov 2003)*.
  - The date underneath the database menu will update automatically to reflect the choices made in the *Entries* and database pull-down menus.
  - To restrict the search to a particular date, select **Other** (located at the bottom of the databases pull-down menu) and pick the date of interest.
  - Once the selection has been made, click **OK** to save the changes then click **OK** again to return to the *Search Setup* window.
  - The *Search Setup* window will reflect the changes made in the *Select Previous CSD Version* window:



- **Note**: a search for entries modified before/after a particular date will return entries added before/ after the particular date.

18.3.2 Searching on entires in a pre-defined hitlist

- Build the query as usual and hit **Search**. Then hit the **Select Subset** button in the *Search Setup* dialogue box.
- Select **Entries in a pre-defined hitlist**.
- Click on the button underneath the words **Restrict on pre-defined hitlists** and select the required hitlist. The following pre-defined hitlists are available:
  - **The best representative of a structure**: The CSD can contain multiple determinations of the same crystal structure, reporting improved structural models, or new experiments performed at different temperatures, pressures, or using different radiations. These multiple inclusions could introduce bias in statistical surveys. Searches can be restricted to just the best representative of each unique polymorph, thus eliminating hits from duplicate crystal structures. 4 best representative subsets are available. These are based on R-factor, the presence of hydrogens, room temperature and lowest temperature determinations.
    Further information can be found here:
    Searching the Cambridge Structural Database for the 'Best' Representative of Each Unique Polymorph. J. van de Streek, *Acta Cryst.*, B62, 567-579, 2006
    [DOI:10.1107/S0108768106019677]
  - Classroom subset: Classroom ConQuest is a version of ConQuest which has been designed for group teaching activities. It has all the functionality of *normal* ConQuest with the limitation that searches can only be done on a subset of entries. The subset of entries can either be the default selection supplied with Classroom ConQuest or one derived by the user

from the main CSD (see USING CONQUEST FOR TEACHING - CLASSROOM CONQUEST, page 151).

- Once a pre-defined hitlist has been selected wait a moment while the subset is loaded (indicated by the progress bar reaching 100%). Then hit **OK.**
- Hit **Start Search** to begin the search. The resulting search will be restricted to only those entries that are part of the pre-defined hitlist.
- Hit **Clear Subset** when the *Search Setup* dialogue box is next opened to revert to searching on the complete database.

18.3.3 Searching Hits from a Previous Search in the Current ConQuest Session

- Build the query as usual and hit **Search**. Then hit the **Select Subset** button in the *Search Setup* dialogue box.
- Select **Entries in a hitlist loaded this session**.
- Click on the button underneath the words **Restrict based on search this session** and select the required search. (If you are unsure which is which, you will need to go back to *View Results* and look at the search results.) (see Switching between Searches, page 40)
- Wait a moment while the subset is loaded (indicated by the progress bar reaching 100%). The hit **OK.**
- Hit **Start Search** to begin the search. The resulting search will find all the entries from the previous search that satisfy your new query.
- Hit **Clear Subset** when the *Search Setup* dialogue box is next opened to revert to searching on the complete database.

18.3.4 Searching the Hits from a Search Previously Saved to Disk

- Build the query as usual and hit **Search**. Then hit the **Select Subset** button in the *Search Setup* dialogue box.
- Select either **Entries in saved search file** or **Entries in saved refcode list,** depending on whether the subset you wish to search was saved in native ConQuest (*.cqs*) format (see Saving and Reading Searches, page 41) or as a list of refcodes (*.gcd*), respectively (see Saving Individual Structures and Lists of Structures, page 41).
- Click on the **Browse** button and select the required file
- Wait while the subset is loaded (indicated by the progress bar reaching 100%). This may take some time.
- Hit **OK** when the subset is loaded.
- Hit **Start Search** to begin the search. The resulting search will find all the entries in the saved search file or refcode list that satisfy your new query.
- Hit **Clear Subset** when the *Search Setup* dialogue box is next opened to revert to searching on the complete database.

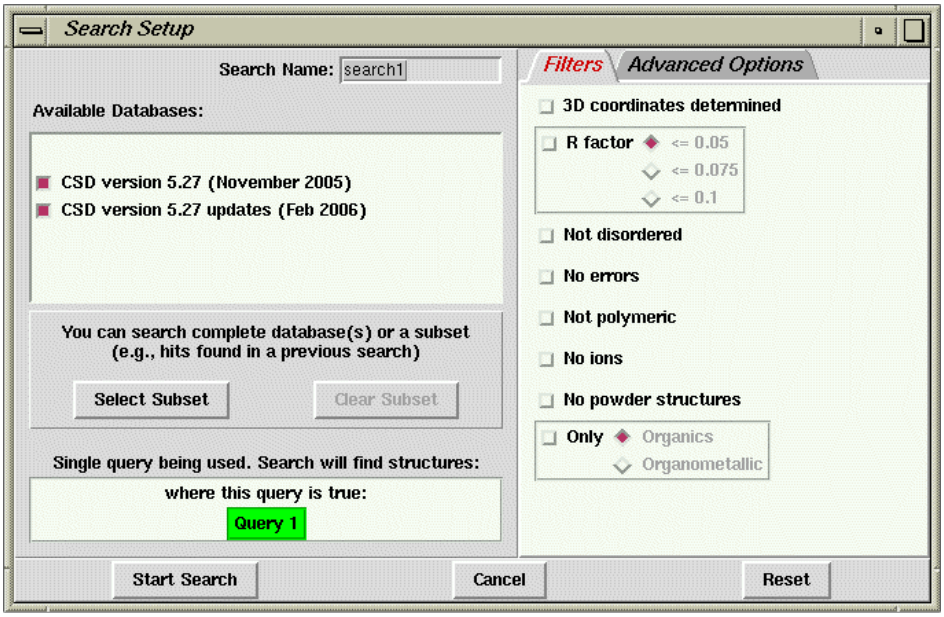
18.4 Searching CSD Data Updates

From April 2002 it has been possible to download CSD data updates at regular intervals. This will keep your copy of the CSD more current between each major release of the CSD System.

The following URL will take you to the download page which describes what you need to do:

http://www.ccdc.cam.ac.uk/products/csd_system/conquest/csd_updates/

- You will be required to enter your Site Code and Confirmation Code in order to download an update.
- The updates will need to be placed in the same directory as your CSD database files (for more details see CSDS Release&Installation Notes).
- Each update will be shown separately in the *View Databases* menu.
- The update packages can be searched either with the main database or separately as desired; this is controlled via options in the *Search Setup* dialogue box:



19 USING CONQUEST FOR TEACHING - CLASSROOM CONQUEST

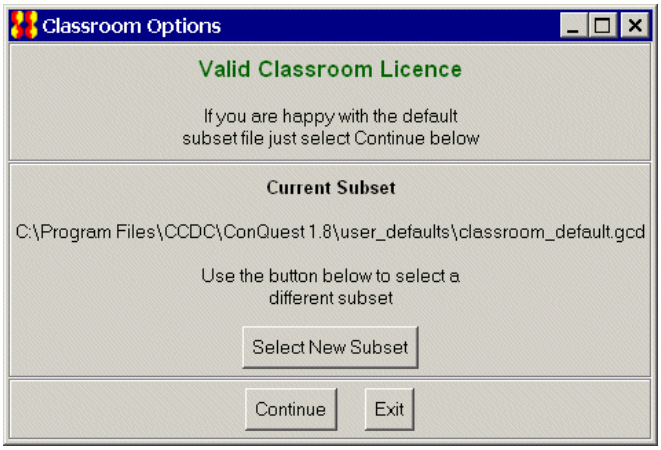
19.1 Main Features of Classroom ConQuest

Classroom ConQuest is a version of ConQuest which has been designed for group teaching activities:

- Anyone with at least one normal ConQuest licence can install as many copies of Classroom ConQuest as they require.
- It has all the functionality of *normal* ConQuest with the limitation that searches can only be done on a subset of entries.
- The subset of entries can either be the default selection supplied with Classroom ConQuest or one derived by the user from the main CSD.

19.2 Basics of Using Classroom ConQuest

Each time a *classroom registered* version of ConQuest is started it will produce a dialogue box similar to the following:



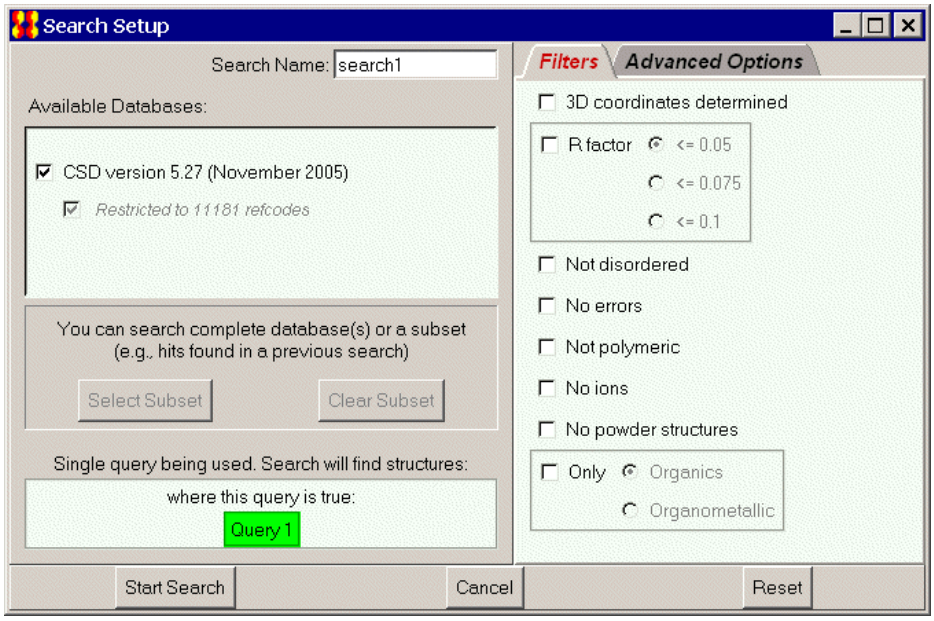
The aim of this dialogue box is to:

- Tell you when the Classroom licence has expired.
- Allow the selection of a different subset for the duration of the session (see Choosing a Subset, page 152).

Within ConQuest everything is normal except for the *Search Setup* dialogue box, which indicates that:

- Only the main CSD is available.
- Searches are restricted to a subset of refcodes (the associated check-button is disabled so that it cannot be changed).

- The **Set Subset** and **Clear Subset** buttons are inactive.



19.3 Choosing a Subset

The subset used by Classroom ConQuest is specified by a file containing a list of CSD refcodes.

The default refcode list is `classroom_default.gcd` which is located in the `user_defaults` directory.

You can elect to use a different subset containing up to one fifth of the current size of the entire database. This can be done by:

- Replacing the default file with a new one that has the same name.
- Editing `user_defaults/conquest_options` so that the line beginning `classroom_refcode_list = ...` indicates the location of the refcode list to be used (specify the full path). For PC Windows use forward slashes (/) in the path instead of back slashes (\).
- Selecting a different subset using the **Select New Subset** button to select a different refcode list. This button is displayed when each Classroom ConQuest session is started (see Basics of Using Classroom ConQuest, page 151).

Although Classroom ConQuest will start if your refcode list is too big, it will issue an error message when you attempt to start a search.

20 RUNNING CONQUEST IN BATCH MODE

- ConQuest can be run from the command line using the following commands:
  - On Windows platforms: `<CQDIR>\exe\cqbatch.exe`
  - On UNIX platforms: `<CQDIR>/bin/cqbatch`
- Any queries or problems should be directed to *support@ccdc.cam.ac.uk*.
- Command line ConQuest is referred to as *CQBatch* throughout this section.

20.1 Input Query Files

- The following queries can be run using CQBatch:
  - query (*.qry*) or search (*.cqs*) files. These are generated and saved via ConQuest. Manual editing of either file type is discouraged as it may cause CQBatch to behave in an unexpected way.
  - a QUEST instruction file (*.con* or *.que*), created externally to ConQuest. Information about QUEST instruction files can be found in the QUEST documentation: *http://www.ccdc.cam.ac.uk/support/documentation/quest/volume1/z1c02009.html#2_5*.
    Note: not all QUEST instruction file commands may be recognised.
  - MDL MOL files (*.mol*) created externally to ConQuest. These may contain geometric parameters (see Bibliographic Information in the CSD, page 8).

- Input files can be provided to CQBatch via the following combinations of command line options:
  - `-rerun <file>` where `<file>` is either a *.cqs*, *.qry*, *.mol* or QUEST instruction file
  - one or more of `-require`, `-not`, `-either` with each followed by one or more *.qry* or *.mol* files
- A jobname must be provided using the `-j <jobname>` command.
- *Note*: `-rerun` cannot be used with `-require`, `-not`, `-either`.

- To run a straightforward query using CQBatch (assuming the correct environment variables and aliases are set), the following command would be used (where `structure.qry` is the query):

```
cqbatch -j <jobname> -rerun structure.qry
```

20.2 Logic

- Boolean logic can be applied to searches run using CQBatch. The following commands correspond to the *AND*, *NOT* and *OR* options available in ConQuest's *Combine Queries* pane (see Combining Queries, page 20).
  - `-require` will perform a boolean *AND* search using the specified queries, e.g. the command

`cqbatch -j <jobname> -require <query1_file_name> <query2_file_name>`
will retrieve all CSD entries that match *query1* and *query2*.

- `-not` will perform a boolean *NOT* search using the specified queries, e.g. the command
  `cqbatch -j <jobname> -not <query1_file_name> <query2_file_name>`
  will retrieve all CSD entries that do not match *query1* and *query2*.
- `-either` will perform a boolean *OR* search using the specified queries, e.g. the command
  `cqbatch -j <jobname> -either <query1_file_name> <query2_file_name>`
  will retrieve all CSD entries that match either *query1* or *query2*.

- The above commands can be combined e.g.
  `cqbatch -j <jobname> -require <query1_file_name> -not <query2_file_name> -either <query3_file_name> <query3_file_name>`
  will retrieve all CSD entries matching *query1* and either *query3* or *query4*, which do not match *query2*.
- If a *.qry* file has more than one query in it then these will be treated as follows:
  - `-require`, `-rerun`: the search will find entries that match all queries
  - `-not`: search will find all entries that match none of the queries
  - `-either`: search will find entries that match at least one of the queries
- If `-rerun` is used with *.cqs* files then the logic from the *.cqs* file will be used
- If `-rerun` is used with a QUEST instruction file then:
  - if a QUEST instruction line is present in the file, this will be used to dictate the logic
  - if no QUEST instruction line is found, *AND* boolean logic will be applied to the queries (or tests) in the file i.e. the search will retrieve entries that match all queries.

20.3 Databases

- CQBatch will search all databases found in the main CSD directory by default.
- If this cannot be located, the location must be provided via the GUI or through a number of CQBatch command line options:

none		All database files in the default CSD directory will be searched
`-db`	One or more *database_file*	Full path to one or more database files. Only these will be searched if this is the only option specified. CSD will add CSD and UPDATES will add updates if these can be found.

`-dbdir`	One or more *database_directory*	All databases found in the directories specified will be searched along with any specified by `-db` or `-dblist` (assuming `-dbfilter` does not dictate otherwise).
`-dblist`	One or more *database_list_file*	A *database_list_file* should contain a list of database paths (one per line). If this is the only option specified then only the databases listed will be searched.
`-dbfilter`	One or more *database_names*	Will filter the list of databases defined by other database options (including *none*) so that it only includes those that match the names listed. `CSD` will keep CSD and `UPDATES` will keep updates if these are present in the list defined by other options
`-allcsd`	Force previous versions of the CSD to be included	By default, ConQuest will not include versions of the CSD that are older than the current version unless: • the current version is not specified by other database-related options (in which case all old versions are included) • the `old_databases` flag in *.csds_defaults* is set to something other than `0` • this command line option is specified

20.4 Restrictions

- A saved search (*.cqs*), hitlist combination (*.chl*) or refcode list (*.gcd*) can be used to restrict searches to a subset of entries via the `-restrict` command line option. For example if you wanted to run a search of hitlist combination file *comb.chl*, you would use a command of the type:
  - `cqbatch -j <jobname> -rerun <query_filename> -restrict comb.chl`

20.5 Filters

Filters of the type found in the graphical interface version of ConQuest (see Available Filters, page 144) can be activated when using CQBatch.

- To filter on R factor, the `-rfactor` command must be used.
- The `-rfactor` command must be followed one of the following:

`0.05`	Hits with rfactors greater than 5% are filtered during the search (see Available Filters, page 144).
`0.075`	Hits with rfactors greater than 7.5% are filtered during the search (see Available Filters, page 144).
`0.1`	Hits with rfactors greater than 10% are filtered during the search (see Available Filters, page 144).

- For example, to exclude CSD entries with R factors of greater than 7.5% from a search, a command of the following type could be used:
  - `cqbatch -j <jobname> -rerun <query_filename> -rfactor 0.075`

- Any filters other than R factor require the `-exclude` command.
- The `-exclude` command must be followed by one or more of:

`ionic`	Ionic structures are filtered during the search (see Available Filters, page 144).
`errors`	Structures with errors are filtered during the search (see Available Filters, page 144).
`no3d`	Structures without 3D coordinates are filtered during the search (see Available Filters, page 144).
`powder`	Structures determined from powder X-ray diffraction are filtered during the search (see Available Filters, page 144).
`organometallic`	Organometallic structures are filtered during the search (see Available Filters, page 144).
`organic`	Organic structures are filtered during the search (see Available Filters, page 144).
`disorder`	Disordered structures are filtered during the search (see Available Filters, page 144).
`polymeric`	Polymeric structures are filtered during the search (see Available Filters, page 144).

- For example, disordered CSD entries can be excluded from a search using a command of the

type:

  - `cqbatch -j <jobname> -rerun <query_filename> -exclude disorder`

- More than one filter can be applied, for example to exclude ionic CSD entries without 3D coordinates, a command of the following type could be used:
  - `cqbatch -j <jobname> -rerun <query_filename> -exclude ionic no3d`

- The `-rfactor` and `-exclude` commands can be used simultaneously.
- If the source of a query is a *.cqs* file then CQBatch will override filters used for the original search if these are specified via the command line.
- If the source of a query is a QUEST instruction file then CQBatch will attempt to identify any SCREEN instructions and over-ride these as appropriate if filters are specified via the command line.

20.6 Output

- By default, a *.cqs* file is always written to the working directory at the end of a search.
- The name of this file will be <*jobname.cqs*> i.e. the jobname supplied using the `-j` command. The jobname defined after `-j` will also be used as the root for all other output filenames.
- Hit entries will be automatically exported if one or more export format is specified after the `-export` command.
- ConQuest's default export settings are used unless an alternative is given using the -export-opts command. Available options depend on the file type being exported.
- Valid commands are the following:

`-export`	`cif`		
	Exports the search results in CIF (Crystallographic Information File) format (see CIF Format, page 218).		
		`-export-opts`	`structure-only`
	Only the structure, not the additional text and numeric fields contained in CIFs will be exported.		
		`-export-opts`	`include-bonds`
	Connectivity information is written.		
		`-export-opts`	`include-parameters`

	Data associated with geometric parameters defined in the search query, are output.		
`-export`	`endnote`		
	Exports the search results as tab-separated values which can be used to import references into the EndNote software for publishing and managing data.		
`-export`	`res`		
	Exports the search results as a simulated SHELX *.res* file.		
`-export`	`mol2`		
	Exports the search results in *.mol2* format: Tripos Inc. format for 3D molecules and crystal structures (see Mol2 Format, page 233)		
		`-export-opts`	`ccdc-bondtypes`
	CCDC-specific representations of bond types not recognised by the *.mol2* format are output.		
		`-export-opts`	`normalise-h`
	Distances of C- N- and O-H bonds are normalised using the default H-Normalise settings.		
		`-export-opts`	`largest-molecule-only`
	Exports the largest chemcial unit (i.e. molecule) in the structure only		
`-export`	`cor`		
	Exports the search results in a simple ASCII listing of 3D coordinates (see COORD Format, page 221).		
		`-export-opts`	`hit-fragment`
	Exports only the coordinates for atoms in the hit fragment (i.e. the chemical substructure used the define the search).		
		`-export-opts`	`orthogonal`
	Exports orthogonal coordinates.		
`-export`	`cmf`		
	Exports the search results in CIFMIF format (see CIFMIF Format, page 219).		

		-export-opts	include-parameters
	Data associated with geometric parameters defined in the search query, are output.		
-export	sd		
	Exports the search results in sd format: MDL Information Systems Inc. format for 3D molecules (see SD Format, page 240).		
		-export-opts	single-not-any
	Outputs single bond types for CCDC bond types not recognised by the MOL/SDfile format.		
		-export-opts	normalise-h
	Distances of C- N- and O-H bonds are normalised using the default H-Normalise settings.		
		-export-opts	include-disordered-atoms
	Exports all atoms, including those identified by the CSD editors as being disordered.		
		-export-opts	largest-molecule-only
	Exports the largest chemcial unit (i.e. molecule) in the structure only		
-export	fdat		
	Exports the search results in CCDC's fdat format; necessary for transferring structures into RPluto: *http://www.ccdc.cam.ac.uk/free_services/rpluto/*.		
-export	gcd		
	Exports the search results in a simple ASCII refcode list (see Refcode (entry ID), page 238).		
-export	pdb		
	Exports the search results in *.pdb* format: the Protein Data Bank format for 3D molecules (see Protein Data Bank, page 237)		
-export	txt		

	Exports the search results in plain text format. Use the `-export-opts` below to control which data are written to the text file.		
	`tab`		
	Exports the search results in *tab separated value* (*.tsv*) format. use the `-export-opts` below to control which data are written to the text file.		
		`-export-opts`	`bibliographic`
	Publication details for each CSD entry in the results hitlist are written to the *.tsv* or *.txt* file.		
		`-export-opts`	`chemical`
	Chemical data (e.g. compound name, formula) for each CSD entry in the results hitlist are written to the *.tsv* or *.txt* file.		
		`-export-opts`	`experimental`
	Details of the structure determination (such as R factor, density) are written to the *.tsv* or *.txt* file.		
		`-export-opts`	`crystallographic`
	Crystallographic information (e.g. cell parameters, space group) is written to the *.tsv* or *.txt* file.		
		`-export-opts`	`internal`
	Data of internal relevance to CCDC (e.g. accession date, additional notes) are written to the *.tsv* or *.txt* file.		
		`-export-opts`	`include-disordered-atoms`
	Exports all atoms, including those identified by the CSD editors as being disordered.		
		`-export-opts`	`largest-molecule-only`
	Exports the largest chemcial unit (i.e. molecule) in the structure only		
`-export-params`	`tsv`		
	Exports geometric parameter data as tab seperated values to the file `<jobname>.tsv`		

	csv		
	Exports geometric parameter data as comma seperated values to the file <jobname>.csv		
	vista		
	Exports geometric parameter data as VISTA files to <jobname>.tab, <jobname>.fgd, <jobname>.fgn		
	sum		
	Exports geometric parameter data as a QUEST summary file to <jobname>.sum		

- The above options correspond to the export formats available in GUI ConQuest (see Saving Individual Structures and Lists of Structures, page 41).
- For *.tab* and *.txt*, the default is to output *bibliographic*, *chemical*, *experimental* and *crystallographic* data unless one or more of `bibliographic`, `chemical`, `experimental` or `crystallographic` are specified via `-export-opts` when only those categories listed will be exported. Specifying `internal` will add CCDC internal data items to those being exported.
- Not all options apply to all formats: if an option is not relevant to a particular format it will be ignored.
- If **SAVE BCCAB** is included in a QUEST instruction file then CQBatch will export a file of hit entries in BCCAB format at the end of the search. This option is provided for those maintaining in-house databases created using PreQuest.
- Progress output will be written to the console window being used to run CQBatch as a search progresses unless redirected to another file via the `-log` option.

20.7 Available Commands

- Some commonly used commands and their usage are given below. Further options can be found by typing `cqbatch -help`.
- Other commands are outlined in other sections:
  - Logic options (see Logic, page 153).
  - Database options (see Databases, page 154).
  - Filter options (see Filters, page 155).
  - Export options (see Output, page 157).

cqbatch	-help	
	Provides guidance about how to use CQBatch and exits.	
	-version	
	Shows the ConQuest and CSD version numbers and exits.	
	-log	<log_filname>
	This command writes diagnostic output to a log file in the directory the search is being run from.	
	-j	<jobname>
	The CQBatch run requires a name <jobname> i.e. the *.cqs* output file, plus any other output files generated from the run are assigned the name <jobname>.	
	-rerun	<query_filename(s)>
	Use this command to re-run a search or query. This option cannot be used with -require, -not or -either.	
	-require	<query_filename(s)>
	Boolean *AND* logic is applied to the query files (in the case of *.cqs* files, the command line logic is applied if no logic is stored within the file).	
	-not	<query_filename(s)>
	Boolean *NOT* logic is applied to the query files (in the case of .cqs files, the command line logic is applied if no logic is stored within the file).	
	-either	<query_filename(s)>
	Boolean *OR* logic is applied to the query files (in the case of .cqs files, the command line logic is applied if no logic is stored within the file).	
	-restrict	
	This option allows searches to be restricted to entries in a particular file, e.g. a previously saved refcode list (see Searching Hits from a Previous Search in the Current ConQuest Session, page 149)	
	-query-directory	
	Rerun all valid files in directory	

	-file-filter	
	filter query directory (valid only with -query-directory)	
	-use-as-query	
	Use entries or search results as queries	
	-query-options	
	Options controlling query creation (valid only with -use-as-query)	
	`-stop`	`<n>`
	This option indicates that the search should stop after <n> hits. If <n> equals zero this overrides a stop limit set anywhere else (e.g.in an instruction file or saved search)	
	`-cqs`	`<filename>`
	The *.cqs* output file is assigned the name `<filename>`.	

21 TUTORIALS

The easiest way to learn how to use ConQuest is to do the tutorials:

21.1 Tutorial 1: Basic Substructure Search

Objectives

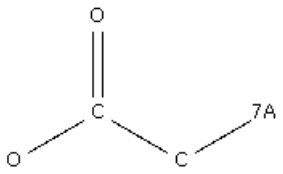
- To draw a chemical fragment (substructure).
- To search the CSD for entries which contain the fragment.
- To obtain a list of journal references for the hit entries.
- Transfer the results to Mercury.

Steps Required

- Draw the fragment in the substructure drawing window.
- Start the search.
- Decide which hit entries to keep.
- Write out the journal references.

The Example

- The tutorial explains how to search for the following fragment:



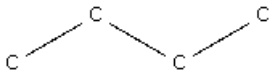
where 7A = F, Cl, Br or I.

Menu Commands Required

1. Start ConQuest and hit the **Draw** button to open the *Draw* window.

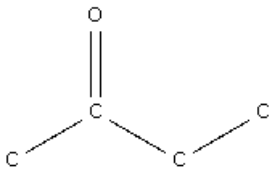
2. Draw a chain of atoms.
   - Move the cursor into the white drawing area, press the left-hand mouse button, and move the

mouse while keeping the button depressed. Release the mouse button and you should have drawn a C-C bond. Keep the cursor where it is, i.e. on one of the carbon atoms. Press the left-hand button again and draw a second bond, and then a third to produce:

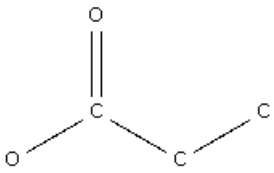


If you make a mistake (here, or elsewhere in the drawing area), click on **Edit** (in the top-level menu) and then **Undo**.

3. Add the carbonyl group.
   - Select O from the list of common elements at the bottom of the *Draw* window.
   - Click on the button next to the word *Bond* at the bottom of the *Draw* window and change the bond type from **Single** to **Double.**
   - Move the cursor back into the white area and draw a bond to the second carbon in the chain. Since the current atom type is O and the current bond type is double, this creates an O=C bond:

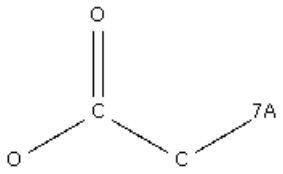


4. Create an acetoxy group.
   - O is the current element, so click on the first carbon in the chain to change it to an oxygen:



5. Create a halogen substituent.
   - Click with the right-hand mouse button on the carbon at the other end of the chain. Select **Element** from the pull-down menu, followed by **More** from the next menu, then **Any Halogen**.

- The terminal carbon should now have changed to 7A (the symbol used in ConQuest to denote the halogen group of elements).



6. The fragment is now drawn. Do the search.
   - Hit the **Search** button in the bottom right-hand corner of the drawing window. This produces a pop-up window telling you which version of the CSD you are going to search.
   - This window allows you to specify some filters. Click on **3D coordinates determined** to ensure that you will only find CSD entries for which 3D atomic positions are available.
   - Hit **Start Search**.

7. Look at the results.
   - As soon as you start the search, the ConQuest interface moves to the *View Results* pane. After a few moments, the refcodes of hit structures start to appear in the list on the right-hand side. Click on any structure to see its chemical diagram.
   - Click on the **Detach Diagram** button, this allows the 2D diagram to be shown in its own window enabling the 2D diagram and any other results panes to be viewed simultaneously. Now use the tabs (*Author/Journal, Chemical,* etc.) to see other types of information.
   - Rather than allowing the search to go to completion, you can stop it by hitting the **Stop Search** button in the bottom right-hand corner. The blue progress bar indicates how much of the CSD has been searched so far. A pop-up menu asks if you really wish to stop the search as it cannot be resumed once stopped.

8. View structures in 3D.
   - Click on the *3D Visualiser* tab to see the currently selected structure in 3D.
   - Rotate the structure by moving the cursor in the 3D-display area while pressing on the left-hand mouse button.
   - Change the display style by clicking anywhere in the 3D-display area with the right-hand mouse button. Select **Display Style** from the pull-down menu, then select a display style.

9. Suppress unwanted hits.
   - Suppose you don't want to keep the first hit. *Suppress* it by clicking on the green tick next to

the refcode. The tick turns into a red cross. This structure will not be written out in any files you export.

10. Writing out a file containing journal references.
    - Click on **File** in the top-level menu and select **Export Entries as...** from the pull-down menu.
    - Click on the file type bar and select **TXT: Text representation** from the pull-down menu**.**
    - Turn the output off for everything except **Bibliographic** information (i.e. click in the check-boxes so that only the one next to **Bibliographic** is coloured red).
    - Hit **Save** to save the file of journal references.

11. Transferring the hits to Mercury.
    - Mercury is an advanced crystal structure visualiser. To view the hits from this search in Mercury hit **Analyse Hitlist** and select **View in Mercury** from the pull-down menu,

This ends the tutorial.

21.2 Tutorial 2: Substructure Search with 2D Constraints

Objectives

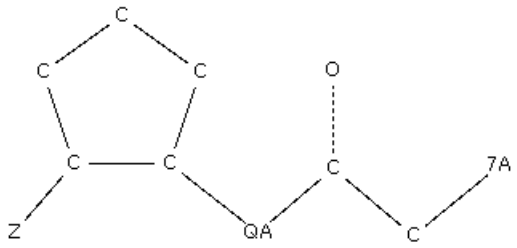
- To search for a chemical fragment, while controlling:
  - The degree of substitution of atoms.
  - The total coordination number (hybridisation state) of atoms.
  - The cyclicity of atoms and bonds (i.e. whether or not they are part of a ring).
  - The chemical environment of the substructure.

Steps Required

- Draw the fragment in the substructure drawing window.
- Specify the *2D constraints*.
- Reject inaccurate structures by selecting only entries with R-factors <= 10% (R-factor is a measure of experimental precision).
- Run the search.
- Save the 3D coordinates of the hit entries.

The Example

- The tutorial explains how to search for the following fragment:

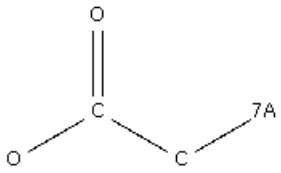


where 7A = F, Cl, Br or I.
QA = C, N or O
Z = any atom except H
QA cannot be part of a ring
Bond C...O can either be single or double

Menu Commands Required

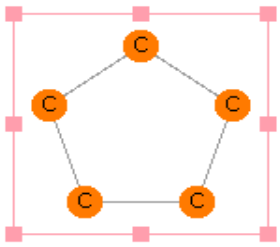
1. Start ConQuest and hit the **Draw** button to open the *Draw* window.

2. Draw the acetoxy structure as shown in Tutorial 1: Basic Substructure Search (see page 164).

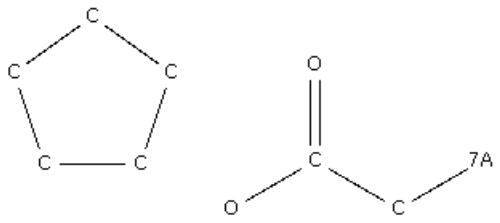


3. Add a cyclopentane ring to the drawing area.
   - Click on the saturated 5-membered ring icon at the left-hand side of the *Draw* window (i.e. the top-left of the four rings). Then click in an empty space in the white drawing area.

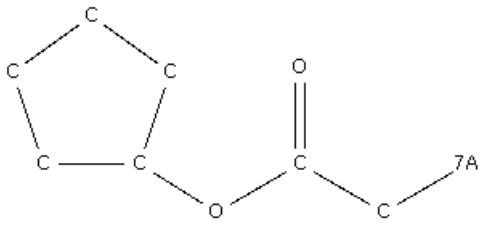
4. Move the cyclopentane ring, if necessary.
   - In a moment, you will make a bond between the O atom of the acetyl group and one of the cyclopentane atoms. Before doing this, you may want to move the ring into roughly the right position to produce a tidy diagram. To do this, click on the **EDIT** button on the left-hand side of the *Draw* window (not the **Edit** button in the top-level menu bar). This puts you into edit mode (previously, you were in draw mode).
   - Select the cyclopentane ring by pressing the left-hand mouse button and dragging the cursor until the ring is enclosed in a rectangle:



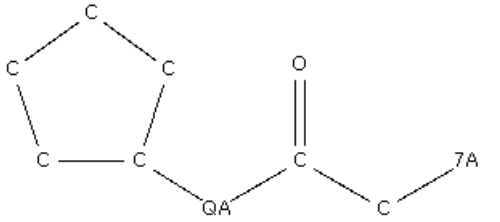
   - Move the ring to its desired position by moving the cursor into the rectangle, pressing down the left-hand mouse button, and *dragging* the ring.



5. Connecting the cyclopentane and acetoxy groups.
   - Go back to draw mode by hitting the **DRAW** button at the left-hand side of the *Draw* window.
   - Position the cursor on the oxygen of the acetoxy group, press down the left-hand mouse button, and move the cursor to one of the ring atoms. This connects the two fragments.

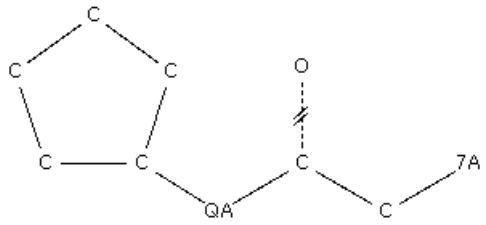


6. Change the linking oxygen to QA = C,N,O.
   - Right-click on the O atom, select **Element** from the pull-down menu, followed by **More** from the next menu, then **Other Elements....**
   - In the Periodic Table window, click on **Multi Pick**, so that you can define a multiple (variable) atom type.
   - Click on the *C, N* and *O* element symbols in the Periodic Table and then hit **OK.** This should change the acetoxy oxygen into a variable atom type *QA*.

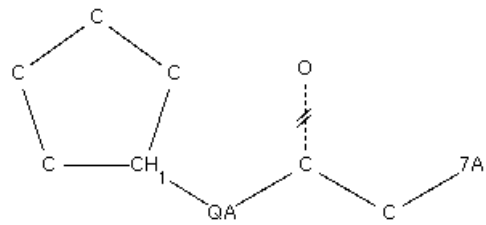


   - Move the cursor onto the *QA* symbol and look at the bottom left-hand corner of the drawing area. There should be a message explaining that QA = C, N or O.

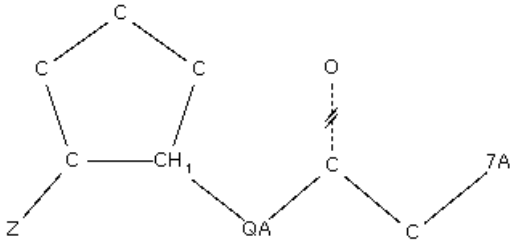
7. Define the C-O bond to be single or double.
   - Right-click on the middle of the C=O bond, select **Type** from the pull-down menu, then **Variable...**
   - In the pop-up window, switch on **Single** and **Double** and hit **OK**.
   - Put the cursor on the middle of the C-O bond and look at the message in the lower left-hand corner which confirms that the bond type can be single or double.



8. Specify CH_1 at the ring substitution atom.
   - Right-click on the ring carbon atom that is bonded to the QA atom, select **Hydrogens** from the pull-down menu, then click on **1** in the next pull-down.

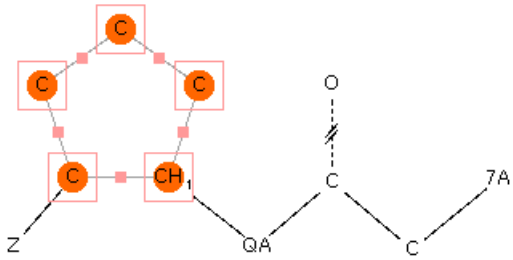


9. Specify that the next ring carbon must be substituted.
   - Click on the button at the bottom labelled **More...**, then **Not Hydrogen.**
   - This sets up a variable atom type, Z, which will match any element except hydrogen. Draw a bond to one of the ring carbons ortho to the substitution atom:

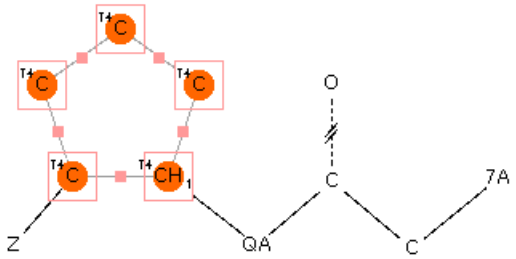


10. Specify sp^3 hybridisation at ring atoms.

- Go into edit mode by clicking on the **EDIT** button at the left-hand side of the *Draw* window (not **Edit** in the top-level menu bar).
- Click and drag to select (i.e. enclose in a rectangle) all the ring carbons.

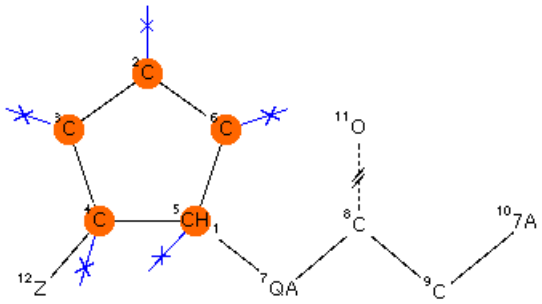


- Select **Atoms** from the top-level menu bar, then **Number of Bonded Atoms**, then **4** from the next pull-down. All the ring carbons are now marked T4, which means that their *total coordination number* must be 4 (i.e. each of the carbons must be bonded to four atoms and hence must be sp^3 hybridised).



11. Specify that QA cannot be part of a ring.
    - Right-click on the *QA* atom, select **Cyclicity**, then **Acyclic.** The atom is now marked with a superscript [a], meaning that it must be acyclic (cannot be part of a ring).

12. Specify that the ring atoms cannot be directly linked to one another.
    - Return to **DRAW** mode.
    - Select **Bonds** from the top-level menu and **Exclude...** from the pull-down menu.
    - Select **Direct Links** and click on all of the atoms in the 5-membered ring.



    - Hit **OK**.

13. Start the search and reject structures with R > 10%.
    - Hit the **Search** button.
    - Switch on the **R factor** filter in the pop-up window, and select **<=0.1.** This will exclude structures with R factors greater than 10%, which tend to be inaccurate.
    - Hit **Start Search**.

14. Save the hits in mol2 format.
    - Stop the search after you have found a few hits. Note that, in each chemical diagram, the search substructure is highlighted in red.
    - Save the 3D coordinates of the hits by selecting **File** from the top-level menu and **Export Entries as...** from the pull-down menu. Set the file type to **MOL2**, make sure **All selected entries** is switched on (to save all the structures, not just the one currently on display) and hit **Save.**

This ends the tutorial.

21.3 Tutorial 3: Defining and Saving Geometric Parameters

Objectives

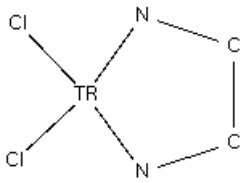
- To draw a chemical fragment and define geometric parameters of interest.
- To search the CSD for entries which contain the fragment.
- To view the geometric parameters for each hit entry.
- To write out a file of the geometric parameters for analysis in Vista or Excel.

Steps Required

- Draw the fragment.
- Define the 3D parameters of interest.
- Run the search.
- Save the parameters in Vista or Excel format.

The Example

- The tutorial explains how to define and save important geometric parameters for the fragment:



where TR = any transition metal and the coordination number of TR is exactly 4. The parameters that will be saved are:

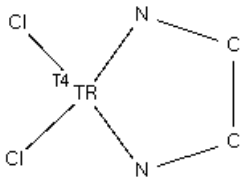
- TR-Cl, TR-N, C-C bond lengths.
- Cl-TR-Cl and N-TR-N bond angles.
- N-C-C-N torsion angle.
- Element symbol of TR.
- R-factor.

Menu Commands Required

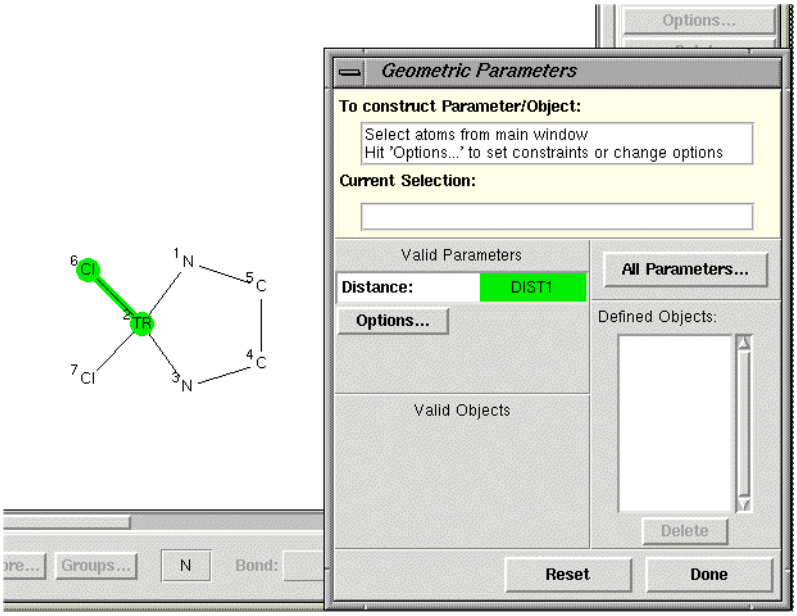
1. Start ConQuest and hit the **Draw** button to open the *Draw* window.

2. Draw the above substructure in the usual way (see Tutorial 1: Basic Substructure Search (see page 164) 1 if you are unsure how to draw substructures).
   - TR (meaning any transition metal) can be selected from the Periodic Table, which is accessible, e.g., by hitting the **More...** button at the bottom of the *Draw* window then **Any Transition Metal.**

3. Specify TR as 4-coordinate.
   - Right-click on TR, select **Number of Bonded Atoms**, then select **4**:



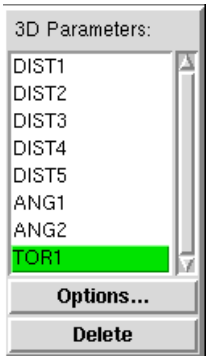
4. Define the geometric parameters.
   - Click on the **ADD 3D** button to the left of the drawing area.
   - Click on one of the Cl atoms and the TR atom, then hit the **Define** button next to the word *Distance* in the *Geometric Parameters* dialogue box:



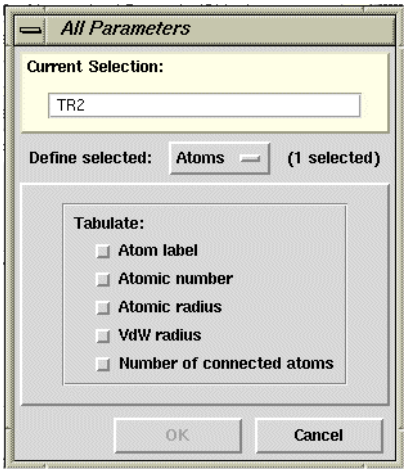
   - This defines a variable, *DIST1*, which will store the value of the TR-Cl bond length. To change the name of the parameter, hit **Options...** and type the name you want into the

RENAME input box. ConQuest puts some restrictions on the name you can use for a parameter. If you type an unacceptable name, the input box will turn pink and you will not be able to hit **OK** to close the pop-up.

- Similarly, define the other Cl-TR distance, the two TR-N distances, and the C-C distance.
- Define the Cl-TR-Cl and N-TR-N angles. Note that, depending on how many atoms you have picked, the *Geometric Parameters* dialogue box only allows you to define meaningful parameters, e.g. if you have picked three atoms, you can define an angle but not a distance or torsion.
- Define the N-C-C-N torsion angle. At this point, all the parameters you have defined will be listed in the *3D Parameters* box in the top right-hand corner of the *Draw* window:



5. Request tabulation of the specific element symbol of TR in each hit.
   - The last parameter to define is the element symbol of TR. Click on TR and then hit the **Define** button next to the words *Atom Properties.* This launches the *All Parameters* window, where a number of atom properties can be selected for output:



   - Click on the **Atom label** radio button, then click **OK** to finish the selection.
   - Tabulating the element symbol is useful because the search will find structures containing a

range of different transition metals. By tabulating the precise element type in each hit structure, it is possible to separate, e.g., Fe from Co from Ni hits in Vista or Excel, thus allowing analysis of differences and similarities between the metals.

- There are no more parameters to define, so close the *Geometric Parameters* pop-up by hitting **Done.**

6. Run the search and view the geometric parameters for specific hits.
   - Hit the **Search** button. Set the filters to exclude polymers, disordered structures, and structures with R-factors greater than 0.1. Also switch on the **No errors** filter. This will exclude structures which contain errors that the editorial staff at CCDC have been unable to resolve.
   - Start the search with the **Start Search** button.
   - Underneath the diagram display for the first entry, there will be a check-box labelled **Show Parameters.** Click on it to display the values, in this particular structure, of the various parameters you have defined.
   - Browse through some of the other hits. Occasionally, you may come across a display something like this:

   - This means that the search fragment occurs more than once in the database structure. As you click on the left- and right-arrows alongside the *Multiple Hits: Show* box, you will see, highlighted in red, the way in which the search fragment is contained within the chemical diagram of the database entry. (If there is no change in the atoms which are highlighted, it means there are two or more crystallographically independent molecules in the structure; the 3D visualiser display will confirm this).

7. Save the geometric data for analysis in Vista or a spreadsheet such as Excel.
   - Depending on whether ConQuest is being run under Unix or Windows, the parameter values may be exported to Vista or a spreadsheet such as Excel for statistical analysis. Select **File** from the top-level menu and then **Export Parameters and Data....** from the resulting pull-down menu.
   - In the resulting pop-up window switch on **R-factor**. This will ensure that the R-factor of each entry is written out as well as the geometrical parameters you have defined.
   - Select either **Vista** (*.tab*) or **Spreadsheet** (*.spr*) as *File type.*
   - Hit **Save** and specify the required file name in the *Save Parameters* dialogue box, then hit **Save** in this dialogue box.
   - The resulting file may then be read into Vista or a spreadsheet package such as Excel,

depending on which type you have saved.

This ends the tutorial.

21.4 Tutorial 4: 3D Substructure Search

Objectives

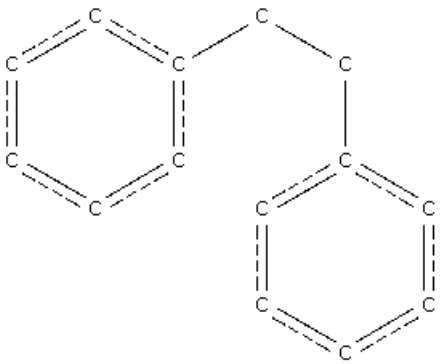
- To draw a chemical fragment (substructure).
- To select only those entries which adopt a particular conformation.
- To tabulate some parameters involving derived objects (centroids, planes).

Steps Required

- Draw the fragment.
- Define centroids and planes.
- Define distances between the centroids, and angles between the planes.
- Define a torsion angle and specify that it must lie in a certain range.
- Run the search.
- Save the file of geometrical parameters.

The Example

- The tutorial explains how to search for the following fragment:



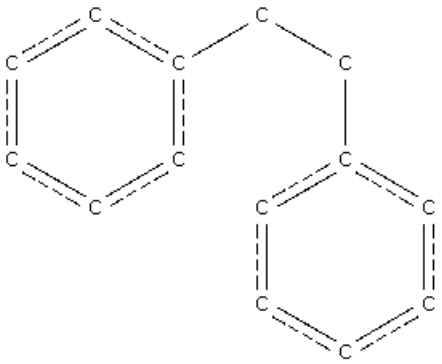
where the two central carbons are acyclic (not part of a ring) and are bonded to 1 or 2 hydrogen atoms. The two phenyl groups will be required to be cis (i.e. C*ring*-C-C-C*ring* torsion angle between –90 and +90 degrees). The following parameters will be saved:

- Distance between the centroids of the phenyl rings.
- Angle between the planes of the phenyl rings.

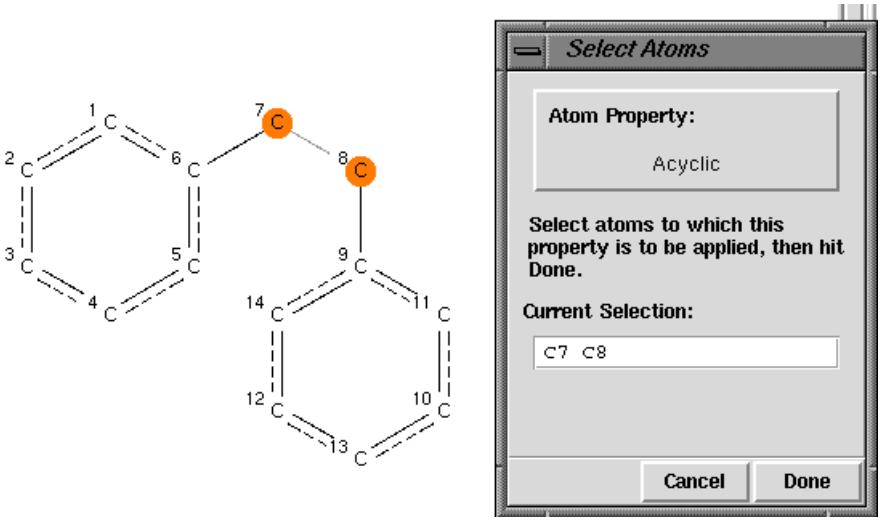
Menu Commands Required

1. Start ConQuest and hit the **Draw** button to open the drawing window.

2. Draw the basic fragment.
   - Use the aromatic bond type for the ring bonds, not alternate double-single (you will get the correct bond types automatically if you select the rings from the ring template area at the bottom left-hand corner of the *Draw* window):

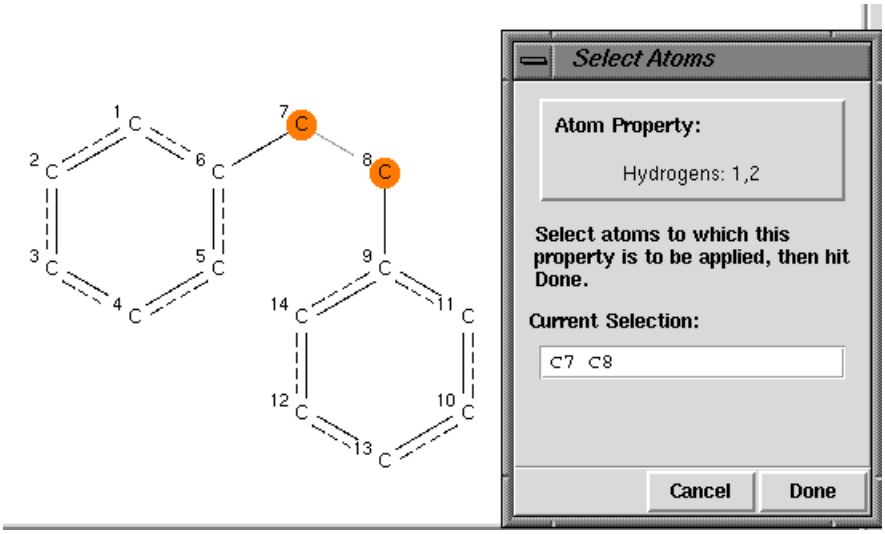


3. Specify the central C atoms to be acyclic.
   - Select **Atoms** from the top-level menu, **Cyclicity** from the pull-down menu, and **Acyclic** from the next menu.
   - Click on the two central carbon atoms:



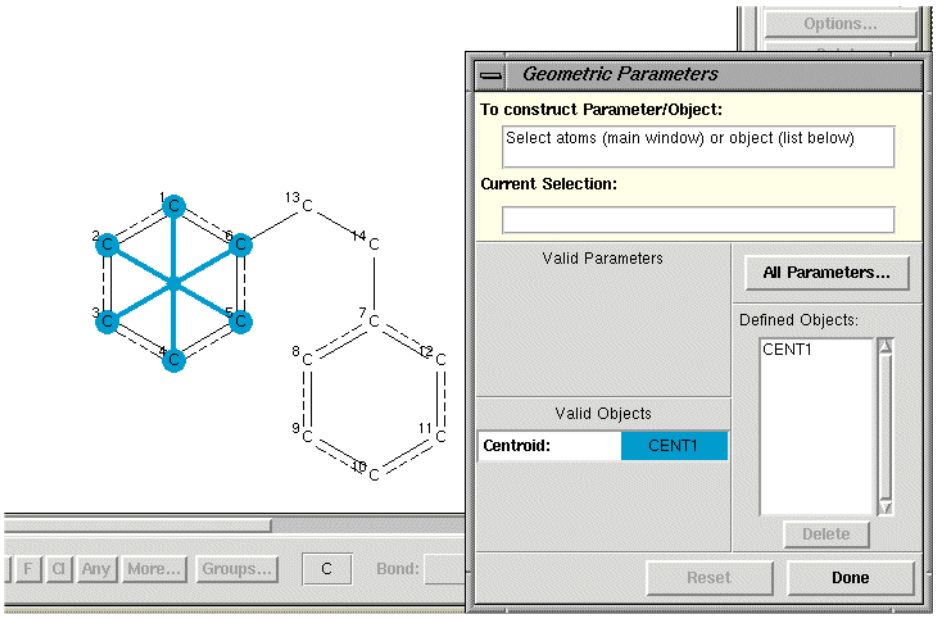
   - Hit **Done** in the *Select Atoms* pop-up.

4. Specify the number of attached hydrogens.
   - Select **Atoms** from the top-level menu, **Hydrogens** from the next menu, and **Other...** from the third menu.
   - Type *1,2* into the *Hydrogen Count* box and hit **OK.**
   - Click on the two central carbon atoms:



   - Hit **Done**.

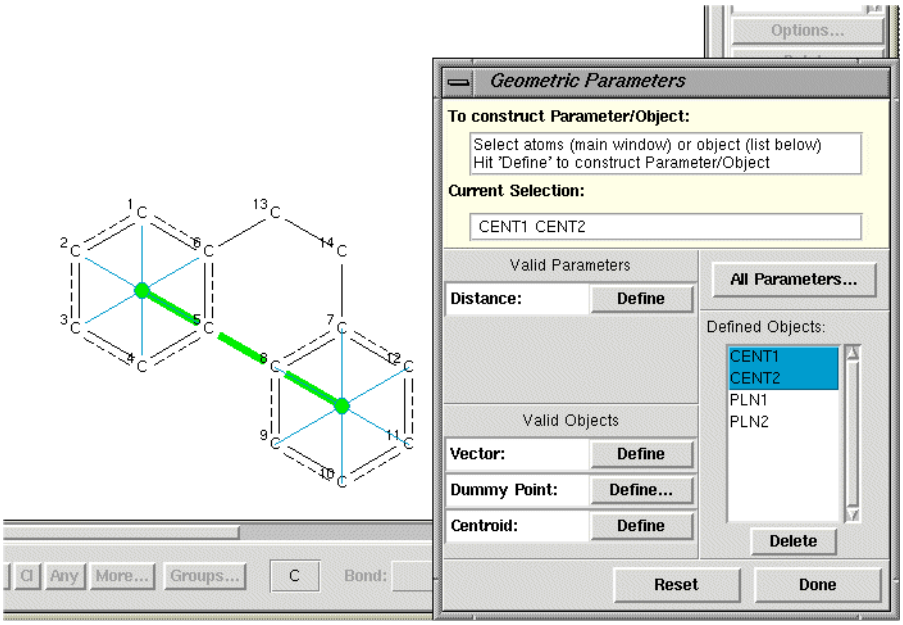
5. Define the centroids of the phenyl rings.
   - Click on the **ADD 3D** button.
   - Click on all six atoms of one of the phenyl rings and hit the **Define** button next to the word *Centroid*. This defines the centroid of the ring (named, by default, *CENT1*):



- Similarly, define the centroid of the other ring (named, by default, *CENT2*).

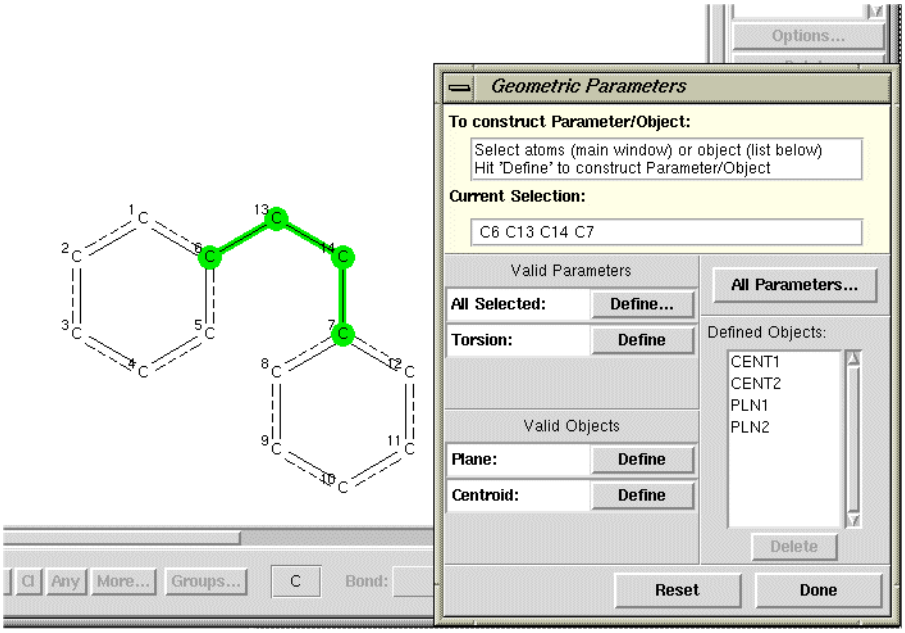
6. Define the ring planes in a similar way.
   - Click on the six atoms of one of the rings and hit the **Define** button next to the word *Plane*. By default, the plane is called *PLN1*.
   - Repeat for the second ring (named, by default, *PLN2*).

7. Define the distance between the centroids and the angle between the planes.
   - Click on the names of the two centroids, *CENT1* and *CENT2*, in the box headed *Defined Objects*:

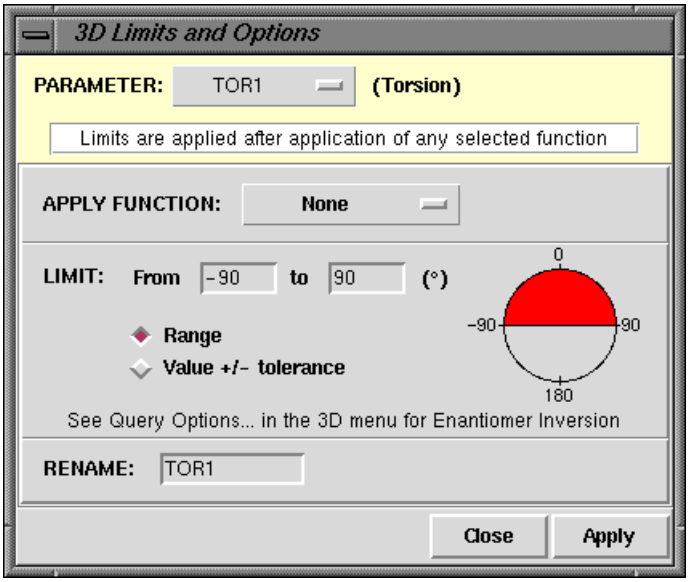


   - Then hit the **Define** button next to the word *Distance*. This specifies a parameter, *DIST1*, which is the distance between the two ring centroids.
   - Similarly, click on *PLN1* and *PLN2* and **Define** the angle between these planes.

8. Constrain the search so that it will only find structures in which the phenyl groups are cis.
   - Click on the four atoms defining the conformation around the central C-C bond:



- Define the C-C-C-C torsion angle by hitting the **Define** button next to the word *Torsion*.
- Hit the **Options...** button.
- Type the required limits *(From* –90 *to* 90) into the *3D Limits and Options* window. The specified torsion angle range is displayed graphically:



- Hit **OK** in this window and then **Done** in the *Geometric Parameters* window.

9. Run the search.
   - Hit **Search**, set the filters **R factor** <= *0.1,* **Not disordered**, **No errors** and **Not polymeric**, and then hit **Start Search**.
   - View some of the hits in the 3D visualiser to confirm that the search is only finding structures in which the phenyl rings are cis.
   - Use the arrows on either side of the boxes labelled *Param* and *Objects* to control whether or not distances, angles, centroid and planes are displayed.

   - A file of the parameters you have defined can be saved by selecting **File** from the top-level menu, then **Export Parameters and Data** from the pull-down menu.

This ends the tutorial.

21.5 Tutorial 5: Search for Nonbonded Interactions

Objectives

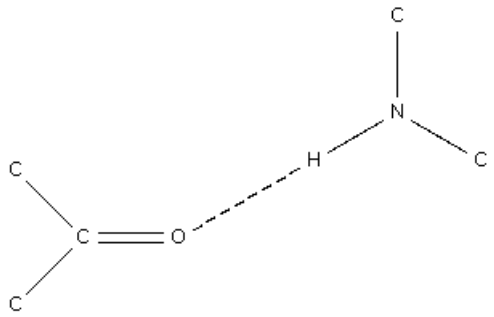
- To draw two chemical fragments.
- To specify that an intermolecular, nonbonded interaction must occur between the fragments.
- To find such nonbonded interactions in the CSD.
- To tabulate some geometrical parameters.

Steps Required

- Draw the fragments.
- Define the required nonbonded contact.
- Define the geometrical parameters that are to be tabulated.
- Run the search.
- Save the file of geometrical parameters.

The Example

- ConQuest can find intramolecular or intermolecular contacts between individual atoms, pre-defined groups of atoms, or between the centroids of pre-defined groups of atoms. The tutorial explains how to search for the following intermolecular contact:



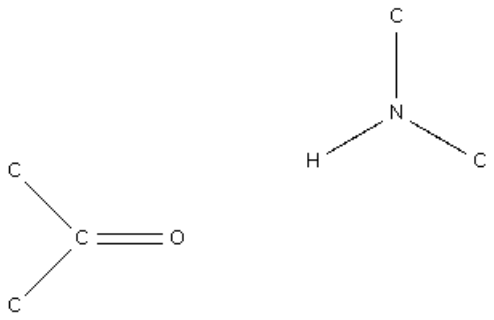
where N is bonded to exactly 3 atoms, and there is an intermolecular contact < 2.1Å between O and H. The following parameters will be saved:

- O...H, O...N nonbonded distances.
- O...H-N and C-O...H angles.
- O and H atom labels.
- Orthogonal coordinates of the hit fragments.

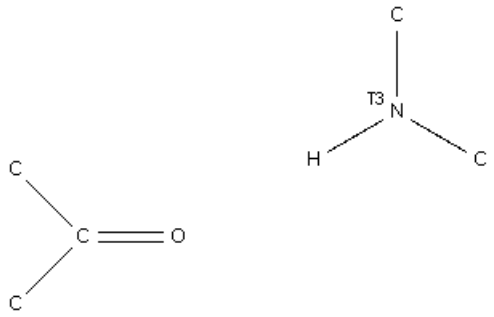
Menu Commands Required

1. Start ConQuest and hit the **Draw** button to open the *Draw* window.

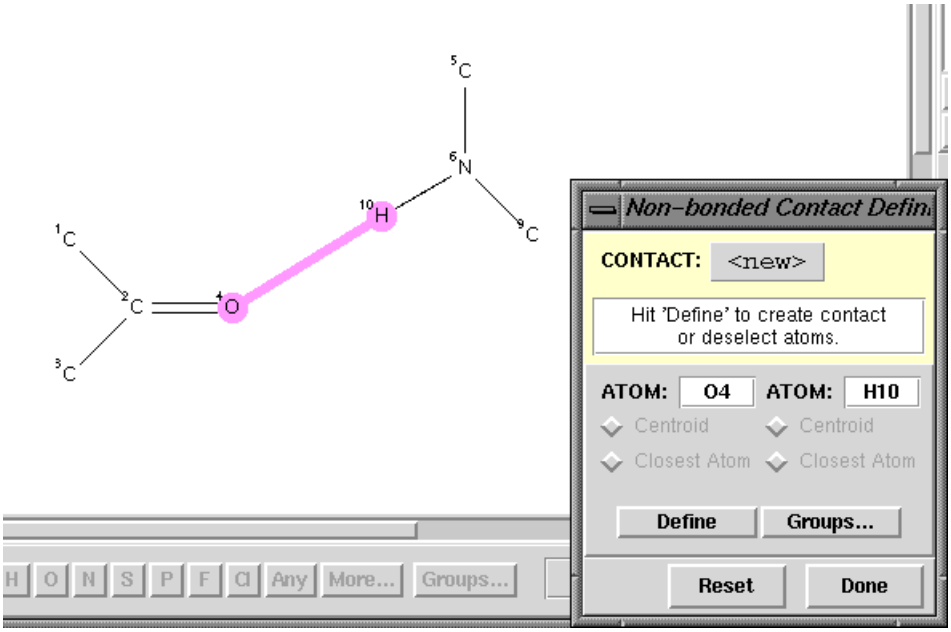
2. Draw the two fragments.



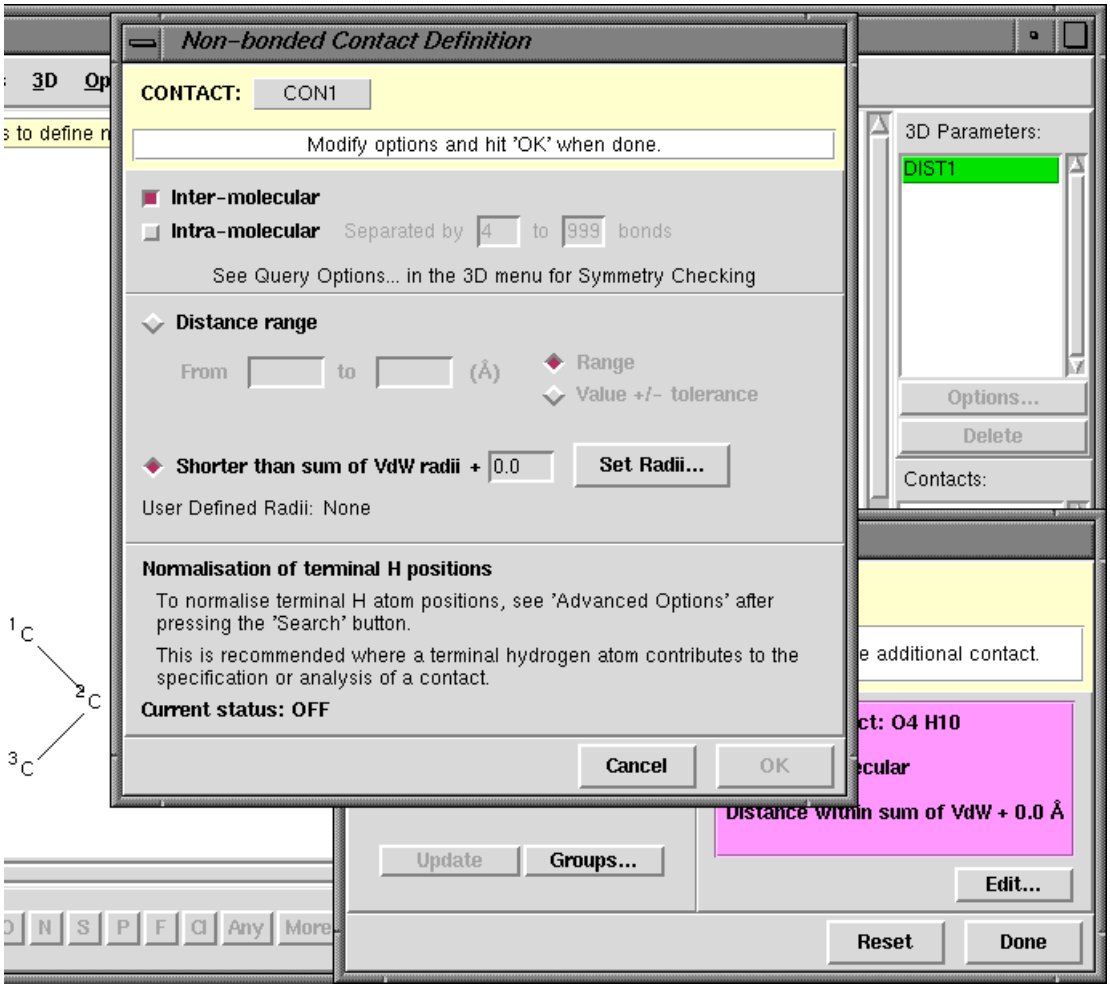
3. Constrain N to be 3-coordinate.
   - Right-click on the N atom, select **Number of Bonded Atoms**, then select **3**.



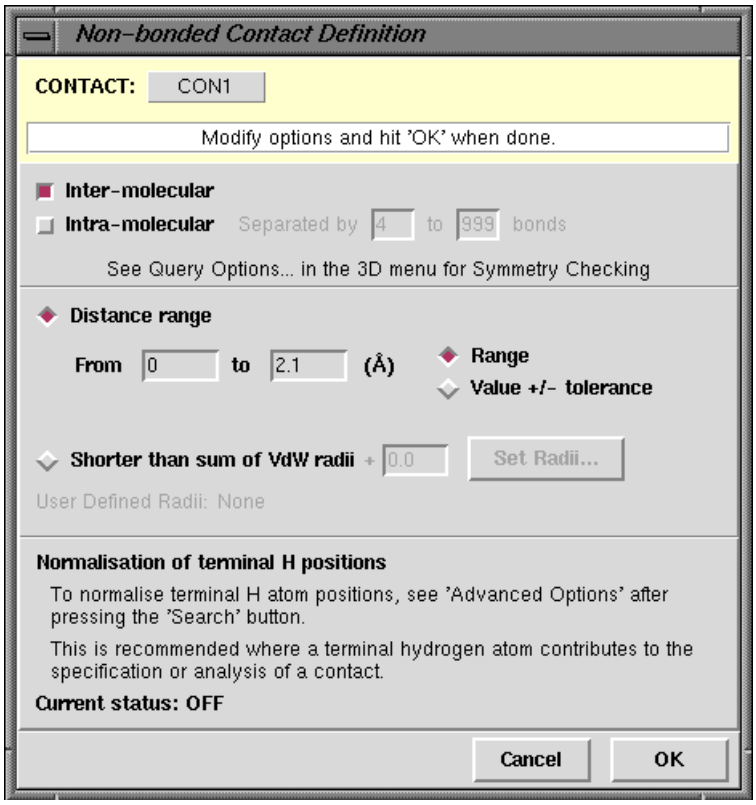
4. Specify nonbonded search criteria.
   - Hit the **CONTACT** button at the left-hand side of the *Draw* window.
   - Click on the O and H atoms and hit the **Define** button:


Non-bonded Contact Defin
CONTACT: <new>
Hit 'Define' to create contact or deselect atoms.
ATOM: O4 ATOM: H10
Centroid Centroid
Closest Atom Closest Atom
Define Groups...
Reset Done
H O N S P F Cl Any More... Groups...

- Click on the **Edit...** button in the *Non-bonded Contact Definition* pop-up:

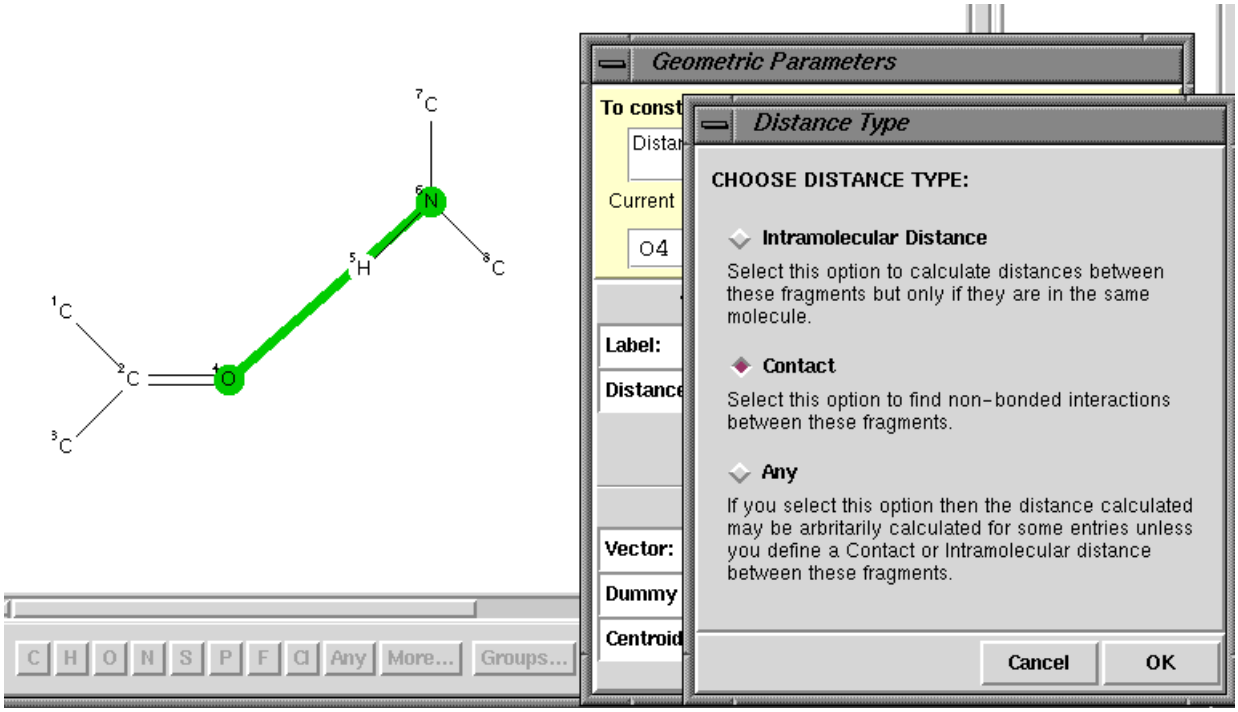

Non-bonded Contact Definition
CONTACT: CON1
Modify options and hit 'OK' when done.
Inter-molecular
Intra-molecular Separated by 4 to 999 bonds
See Query Options... in the 3D menu for Symmetry Checking
Distance range
From to (Å)
Range
Value +/- tolerance
Shorter than sum of VdW radii + 0.0
Set Radii...
User Defined Radii: None
Normalisation of terminal H positions
To normalise terminal H atom positions, see 'Advanced Options' after pressing the 'Search' button.
This is recommended where a terminal hydrogen atom contributes to the specification or analysis of a contact.
Current status: OFF
Cancel OK
3D Parameters:
DIST1
Options...
Delete
Contacts:
Update Groups...
Edit...
Reset Done

- By default, the search is for intermolecular contacts, which is what we want in this example.
- By default, the search will find all contacts shorter than the sum of the van der Waals radii of the interacting atoms. However, in this example, we want to find contacts shorter than 2.1Å. Click on the **Distance Range** radio button, and type *0* and *2.1* into the *From* and *to* boxes, respectively:

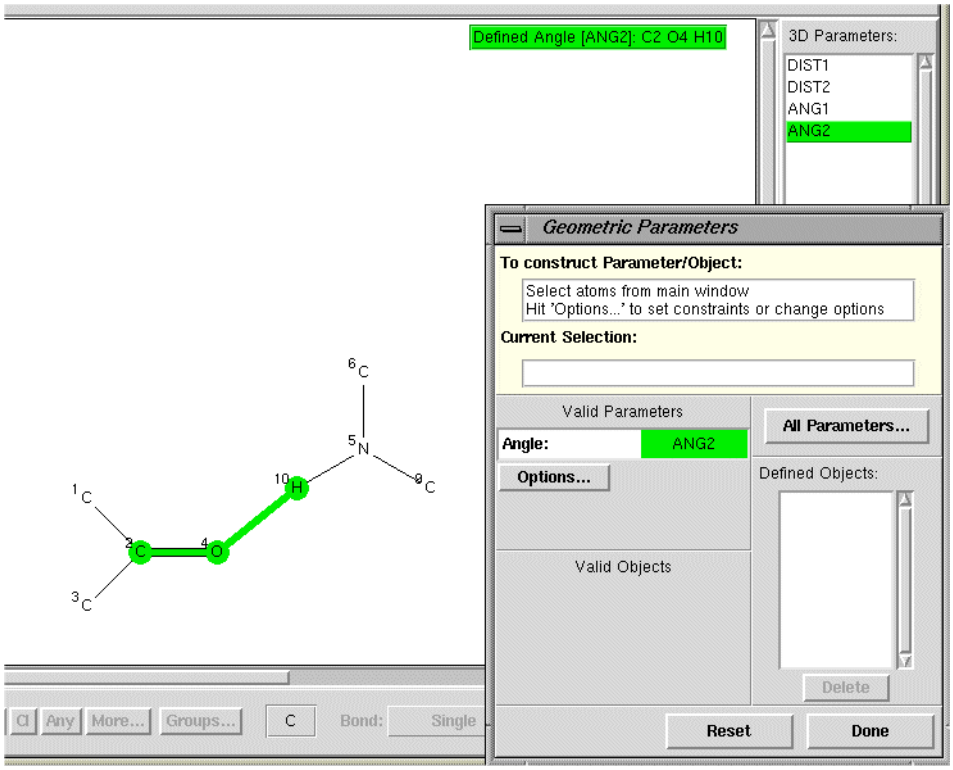


- Hit **OK** and then **Done** to finish definition of the contact.

5. Define the required geometrical parameters and atom labels.
   - Hit the **ADD 3D** button and define the O...N distance as a parameter to be tabulated. This is done by clicking on the O and N atoms and hitting the **Define** button next to the word *Distance* in the *Geometric Parameters* pop-up. Hit **OK** in the *Distance Type* pop-up that then appears to specify that the defined distance is a contact.


Geometric Parameters
To const
Distar
Current
O4
Label:
Distance
Vector:
Dummy
Centroid
C H O N S P F Cl Any More... Groups...
Distance Type
CHOOSE DISTANCE TYPE:
Intramolecular Distance
Select this option to calculate distances between these fragments but only if they are in the same molecule.
Contact
Select this option to find non-bonded interactions between these fragments.
Any
If you select this option then the distance calculated may be arbitrarily calculated for some entries unless you define a Contact or Intramolecular distance between these fragments.
Cancel
OK

- Hit the **Define** button, **OK** and then **Done** in the *Geometric Parameters* dialogue box to finish the definition.
- Also define the O...H-N and C-O...H angles as geometrical parameters.



- Also define the *Labels* of O and H (See Tutorial 3: Defining and Saving Geometric Parameters (see page 174)).

6. Run the search and save the parameters and the coordinates of the hit fragments.
   - Hit **Done,** if you have not already done so, to close down the *Geometric Parameters* window, then hit **Search**.
   - Hit the *Advanced Options* tab and switch on the **Normalise terminal H positions** check-box. This will apply a correction to remove systematic errors in N-H bond lengths determined by X-ray diffraction.
   - Hit the *Filters* tab, select the **R factor** <= *0.1* filter and hit **Start Search**.
   - The interacting atoms are highlighted in the chemical diagram and the hydrogen-bond distances can be displayed by switching on the **Show Parameters** check-box. The intermolecular contact can be viewed in the 3D visualiser.
   - Stop the search when you have a few hits. Select **File** from the top-level menu and then **Export Entries as...** from the pull-down menu. Select the file type as **COORD**.
   - This format enables you to write out either (i) orthogonal or (ii) crystallographic (fractional) atomic coordinates, of either (a) all the atoms in each structure, or (b) just those atoms involved in the contact. Select **Orthogonal** and **Hit Fragment Only** to save the Cartesian coordinates of the atoms involved in the hydrogen-bonding interaction.

This ends the tutorial.

21.6 Tutorial 6: Pharmacophore Search

Objectives

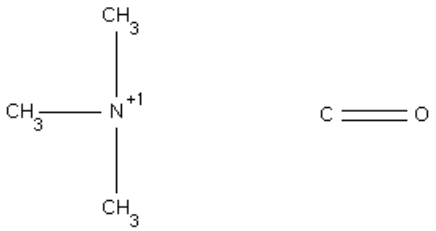
- To draw two chemical fragments (substructures).
- To specify the distance required between the two fragments, i.e. to specify a pharmacophore.
- To find molecules in the CSD containing the specified pharmacophore.

Steps Required

- Draw the fragments.
- Define one of the fragments as a *group* for use in nonbonded searching.
- Specify the required distance between the fragments.
- Ensure that the search is for intramolecular nonbonded distances only (i.e. both fragments must be in the same molecule).
- Run the search.
- Save the file of hits.

The Example

- The tutorial explains how to search for the following pharmacophore:

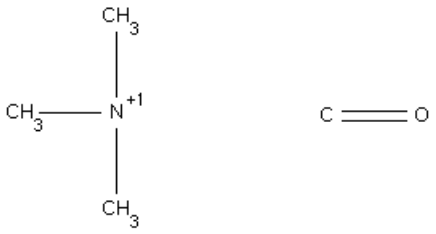


where the distance between the cationic nitrogen and the centre of the carbonyl bond is in the range 5.0 - 5.8Å. This is the distance range required between these two functional groups when acetylcholine and its analogues bind to the enzyme acetylcholinesterase.

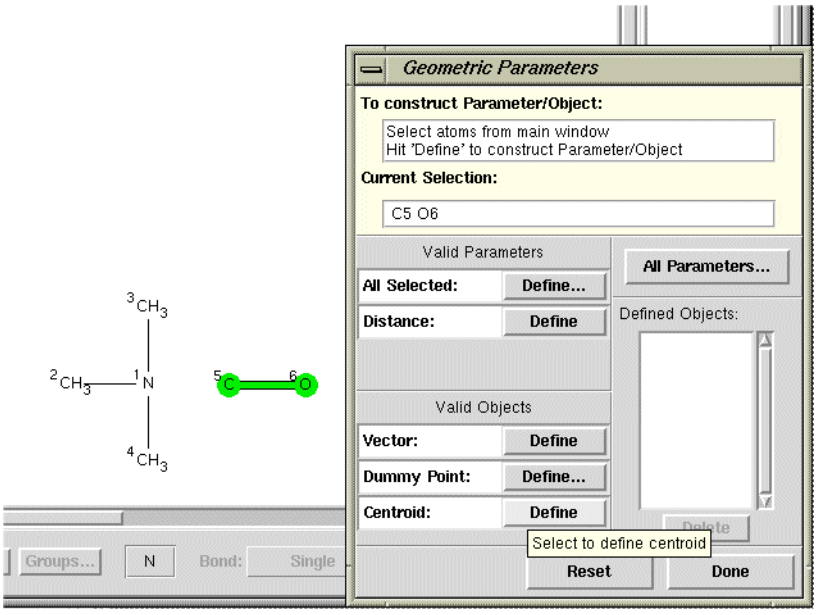
Menu Commands Required

1. Start ConQuest and hit the **Draw** button to open the drawing window.

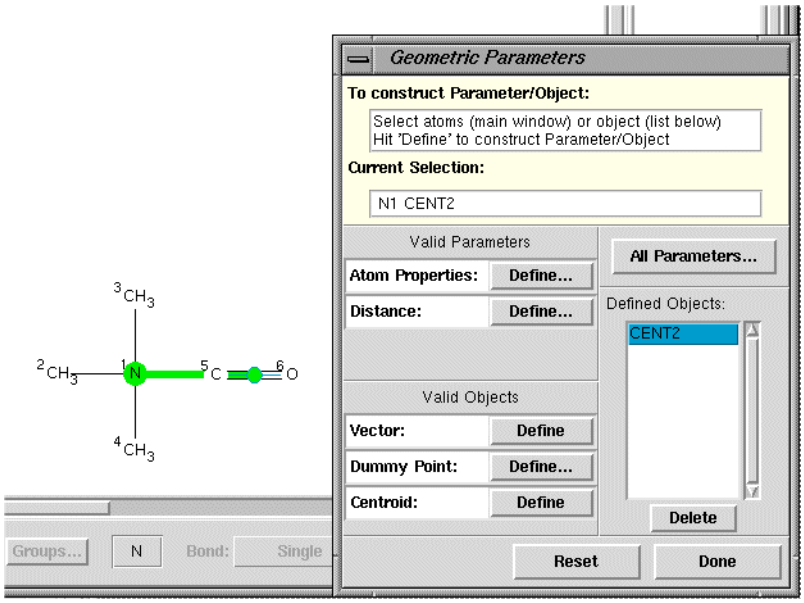
2. Draw the two fragments.
   - Remember to put a charge of +1 on the nitrogen, e.g. by right-clicking on the N atom and selecting **Charge, Positive** and then **+1** from the resulting pull-down menu.
   - If you have problems drawing the substructures, go back and do tutorials 1 and 2.



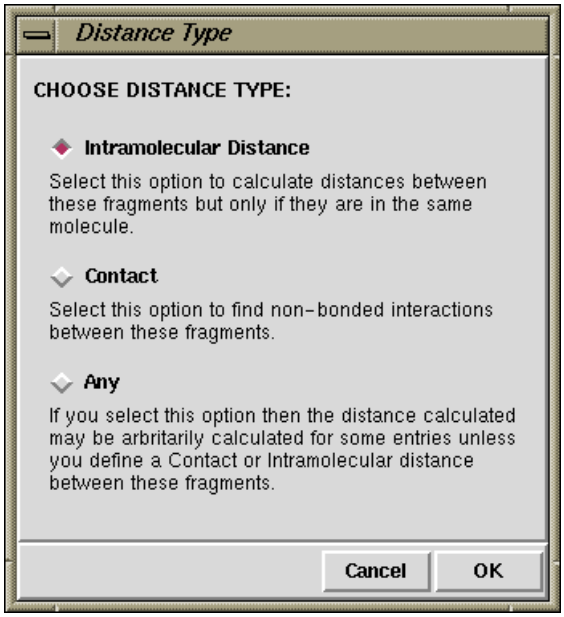
3. Define the centroid of the carbonyl group.
   - Hit the **ADD3D** button at the left-hand side of the *Draw* window.
   - Click on both atoms of the carbonyl group.
   - Click on the button labelled **Define** next to the word *Centroid*:



4. Define the distance between the N atom and the centroid of the C=O group.
   - Select the N atom by clicking on it.
   - Select the centroid of the carbonyl group by clicking on *CENT1* in the box headed *Defined Objects*.
   - Hit the **Define** button next to the word *Distance*:

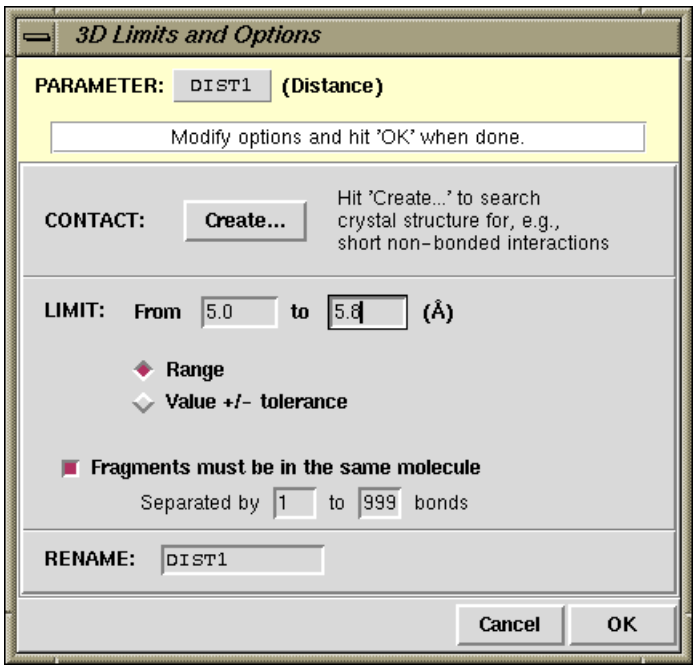


   - Hit **OK** in the resulting pop-up window to specify that the defined distance is intramolecular, i.e. both the trimethylammonium group and the carbonyl group must be in the same molecule:


Distance Type
CHOOSE DISTANCE TYPE:
Intramolecular Distance
Select this option to calculate distances between these fragments but only if they are in the same molecule.
Contact
Select this option to find non-bonded interactions between these fragments.
Any
If you select this option then the distance calculated may be arbitrarily calculated for some entries unless you define a Contact or Intramolecular distance between these fragments.
Cancel
OK

5. Specify the required distance range

- Hit **Options...**
- Type 5.0 in the *From* box and 5.8 in the *to* box:


3D Limits and Options
PARAMETER: DIST1 (Distance)
Modify options and hit 'OK' when done.
CONTACT: Create...
Hit 'Create...' to search crystal structure for, e.g., short non-bonded interactions
LIMIT: From 5.0 to 5.8 (Å)
Range
Value +/- tolerance
Fragments must be in the same molecule
Separated by 1 to 999 bonds
RENAME: DIST1
Cancel
OK

- Hit **OK**.
- Hit **Done** in the *Geometric Parameters* dialogue box to finish definition of the pharmacophore.

6. Run the search.
   - Hit **Search**, set the filters **R factor** <= *0.1,* **Not disordered** and **No errors**, and then hit **Start Search**.
   - View some of the hits, e.g. ACYPAC10, which is a close acetylcholine analogue.
   - You can save the hits (once the search is finished) by selecting **File** from the top-level menu then **Export Entries as...** from the pull-down menu. A variety of formats are available.

This ends the tutorial.

21.7 Tutorial 7: Cell Parameter Search

Objectives

- To find out if the unit cell dimensions of a particular crystal match those of any entries in the CSD. The usual reason for this type of search is to find out whether a crystal of known cell dimensions has already had its structure determined.

Steps Required

- Perform a reduced cell parameter search.

The Example

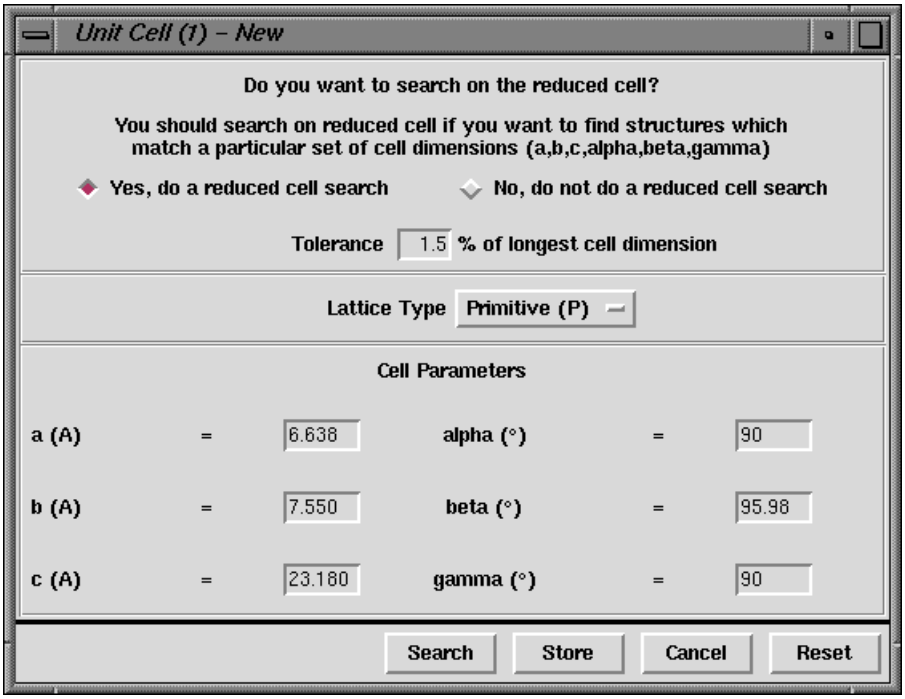
- Find structures whose cell dimensions match those of a primitive monoclinic cell with a = 6.638, b = 7.550, c = 23.180, beta = 95.98.

Menu Commands Required

1. Start ConQuest and hit the **Unit Cell** button.

2. Turn on the option **Yes, do a reduced cell search**.
   - The unit cell of a crystal can be chosen in many different ways (strictly speaking, an infinite number of ways, though not many of the choices will be convenient). In order to assess whether two structures might have the same unit cell dimensions, it is therefore necessary to convert each set of dimensions to a *standard* form. This is called cell reduction.

3. Enter the lattice type and cell dimensions of your crystal.
   - It is essential to specify correctly the lattice type of your crystal. In this example, set the *Lattice Type* to **Primitive.**
   - Type in the values of a, b, c, alpha, beta and gamma. In this example, they are:


Unit Cell (1) - New
Do you want to search on the reduced cell?
You should search on reduced cell if you want to find structures which match a particular set of cell dimensions (a,b,c,alpha,beta,gamma)
Yes, do a reduced cell search
No, do not do a reduced cell search
Tolerance 1.5 % of longest cell dimension
Lattice Type Primitive (P)
Cell Parameters
a (A) = 6.638
alpha (°) = 90
b (A) = 7.550
beta (°) = 95.98
c (A) = 23.180
gamma (°) = 90
Search
Store
Cancel
Reset

- All the cell parameter boxes must be filled in before you can do the search, even though, in this case, the alpha and gamma angles must be 90 by symmetry.
- Accept the default tolerance. This defines how closely the reduced cell dimensions of a database structure must match the reduced cell dimensions of your crystal in order to be considered a match. The default value has been carefully chosen so is usually best left unchanged.

4. Run the search.
   - Hit **Search** in the *Unit Cell* window, then **Start Search** in the next window.
   - Click on the *Crystal* tab in *View Results* to display the cell parameters of hit entries.
   - Two sets of unit cell dimensions are shown for each hit: the cell parameters chosen by the author and the reduced cell parameters to which they correspond. These may or may not be identical, depending on whether the author's choice of cell happened to correspond to the standard reduced cell.
   - Similarly, the reduced cell dimensions of a hit may or may not look like the parameters you typed in, depending on whether *your* choice of cell dimensions happened to be standard. In this example, it was.
   - In comparing the reduced cell dimensions of a database structure with the reduced cell parameters of the input structure, only the cell lengths are compared. This is because the cell angles can be mathematically unstable (i.e. a very small change in input cell dimensions may produce large changes in reduced cell angles; the same problem does not occur for the cell lengths).

- In this example, there are several database entries which match the input cell dimensions, once cell reduction has been performed. Any one of these might therefore be the same structure as that whose cell dimensions were typed in.

This ends the tutorial.

21.8 Tutorial 8: Journal Reference Search

Objectives

- To find a structure from its journal reference.

Steps Required

- Perform a journal search.

The Example

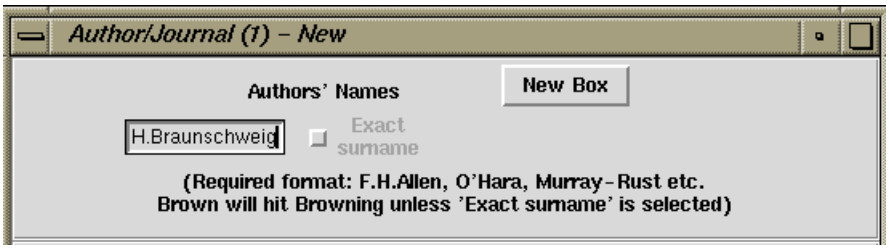
- Find the structure published by H.Braunschweig, P.Paetzold and R.Boese in Chem. Ber., 123, 485 (1990).

Menu Commands Required

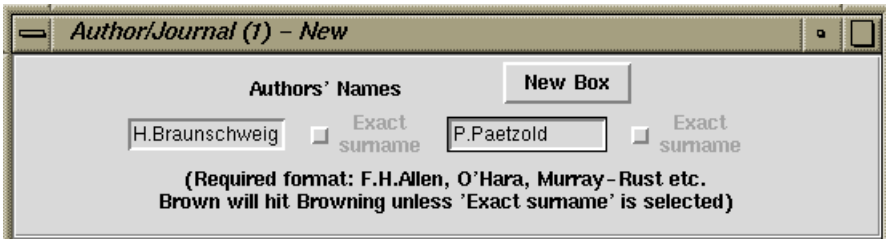
1. Start ConQuest and hit the **Author/Journal** button.

2. Specify the journal reference.
   - It is not necessary to type in all of the authors, journal name, volume number, page number and year – just the journal, volume and page would be enough – but we will do them all for illustrative purposes.

3. Specify the authors' names.
   - Type *H.Braunschweig* into the *Authors' Name*s box; there must be no blank space between the initial and the surname:

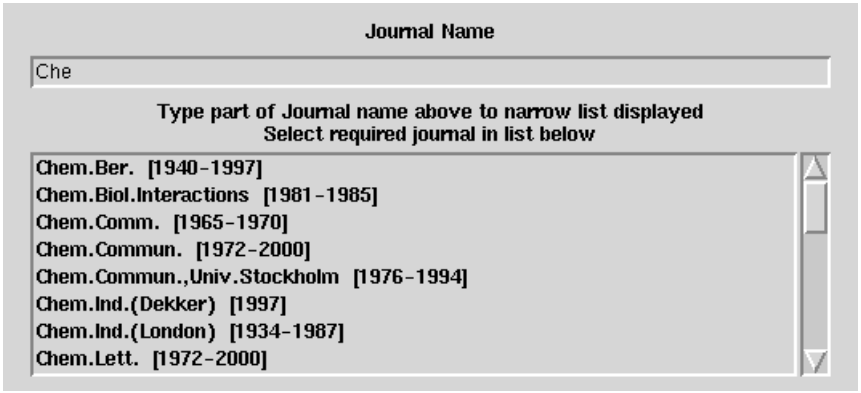


   - Only one name can be typed into each box, so hit the **New Box** button and type in the second author, *P.Paetzold:*



- Hit **New Box** again and type *R.Boese* into the third box.
- Author names are input to the CSD as they appeared in the paper, or in the deposited data if received as a Private Communication.

4. Specify the journal name, volume, page and year.
   - The journal name that you specify must match exactly (i.e. character for character) the name that is stored in the CSD. The safest way of specifying journals is therefore to pick them from the list underneath the *Journal Name* box.
   - Type the three letters *Che* into the *Journal Name* box. The list underneath this box scrolls down to show journals which begin with these three letters:



   - Select *Chem.Ber.* from the list.
   - Type *123* into the *Volume* box.
   - Type *485* into the *Page* box.
   - Type *1990* into the *Year* box.

5. Run the search.
   - Hit **Search** in the *Author/Journal* window, then **Start Search** in the next window.
   - Click on the *Author/Journal* tab in *View Results* to display the literature citation of the hit entry.

This ends the tutorial.

21.9 Tutorial 9: Combined Compound Name and Keyword Search

Objectives

- To create a compound name query.
- To create a keyword query.
- To combine them so as to search for entries containing both the compound name and the keyword.

Steps Required

- Construct and store a compound name query.
- Construct and store a keyword query.
- Select both queries.
- Run the search.

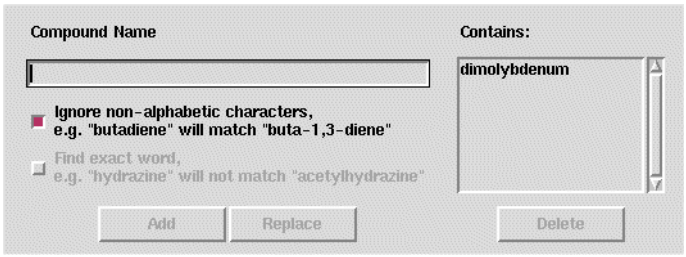
The Example

- Find di-molybdenum structures which are air-sensitive.

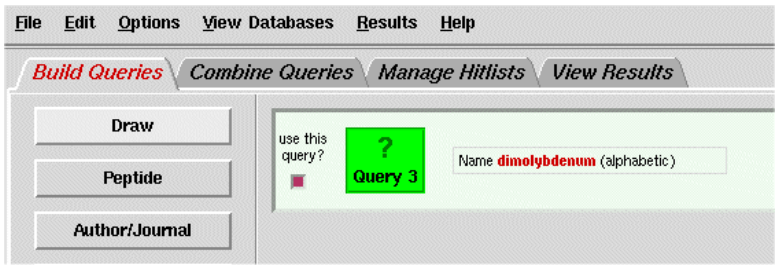
Menu Commands Required

1. Start ConQuest and hit the **Name/Class** button.

2. Specify the required compound name.
   - You don't need to type in a complete compound name, part of a name will do. Type *dimolybdenum* into the *Compound Name* box.
   - You cannot be certain whether names in the CSD contain a hyphen or not, i.e. CCDC editors might have used *dimolybdenum* or *di-molybdenum*. Therefore, click on the check-box **Ignore non-alphabetic character** to ensure that you would find both of these.
   - Click **Add** to add the compound name to the Contains dialogue:

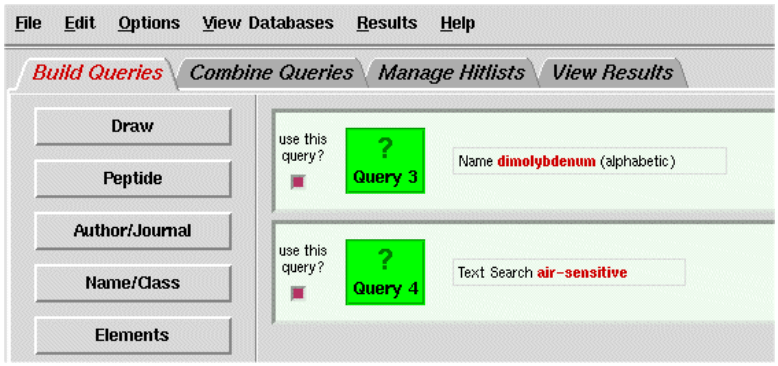


- Hit **Store**. The query is transferred to the *Build Queries* page, but a search is not started:

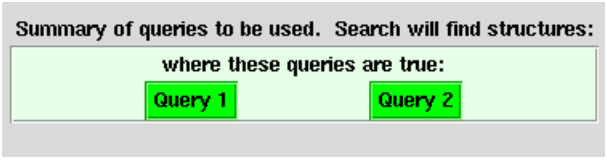


3. Specify the keyword search.
   - Hit the **All Text** button.
   - Select **air-sensitive** from the list.
   - Hit **Store**.

4. Search for both queries.
   - Both of the queries you have built will be coloured green in the *Build Queries* display. This means they are both *selected*.
   - If you click on a query icon, it will turn white which means that it is *de-selected*. If you click on it again, it will turn green and be *selected*.
   - We want to search for entries containing both queries, so, if you have de-selected one or both of them, make sure they are selected again, i.e. both coloured green:



5. Run the search.
   - Hit the **Search** button at the bottom left-hand corner of the window.
   - The pop-up confirms that the search will be for entries that satisfy both queries:



- Hit **Start Search**.
- Click on the *Chemical* tab to show that the hit entries contain *dimolybdenum* or *di-molybdenum* in their compound name, and also contain the words *air-sensitive.*

This ends the tutorial.

21.10Tutorial 10: More Complex Combinations of Queries

Objectives

- To create a formula query.
- To create a space group query.
- To create two queries concerning experimental conditions.
- To combine the four queries in order to find entries that satisfy the formula constraint, do not belong to the specified space group, and satisfy either of the experimental constraints.

Steps Required

- Construct and store a formula query.
- Construct and store a space group query.
- Construct and store two experimental queries.
- Combine the queries (apply Boolean logic).
- Run the search.

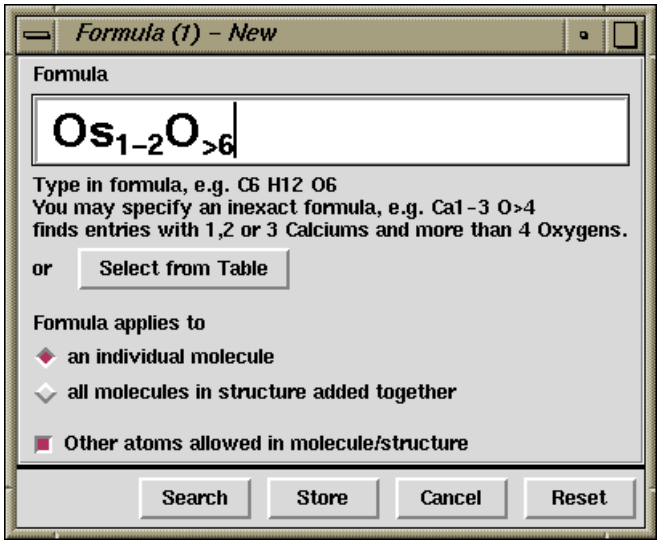
The Example

- Find structures that:
  - contain 1 or 2 osmium atoms and >6 oxygen atoms;
  - do not belong to space group $P2_1/c$;
  - were determined either by neutron diffraction or with an R-factor <= 0.06.

Menu Commands Required

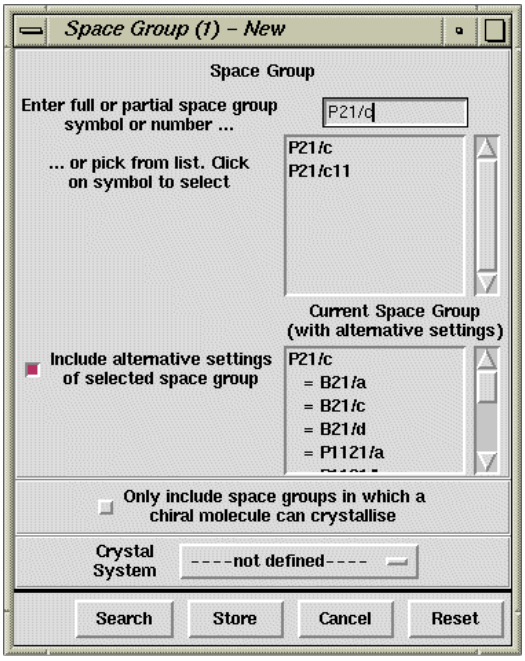
1. Start ConQuest and hit the **Formula** button.

2. Specify that 1-2 Os atoms and > 6 O atoms are required.
   - Click in the white box and type *Os1-2 O>6.*



- Hit **Store** to transfer the query to the *Build Queries* page.

3. Set up a space-group query for P21/c.
   - Hit the **Space Group** button.
   - Type *P21/c* into the space group box or select this symbol from the list:

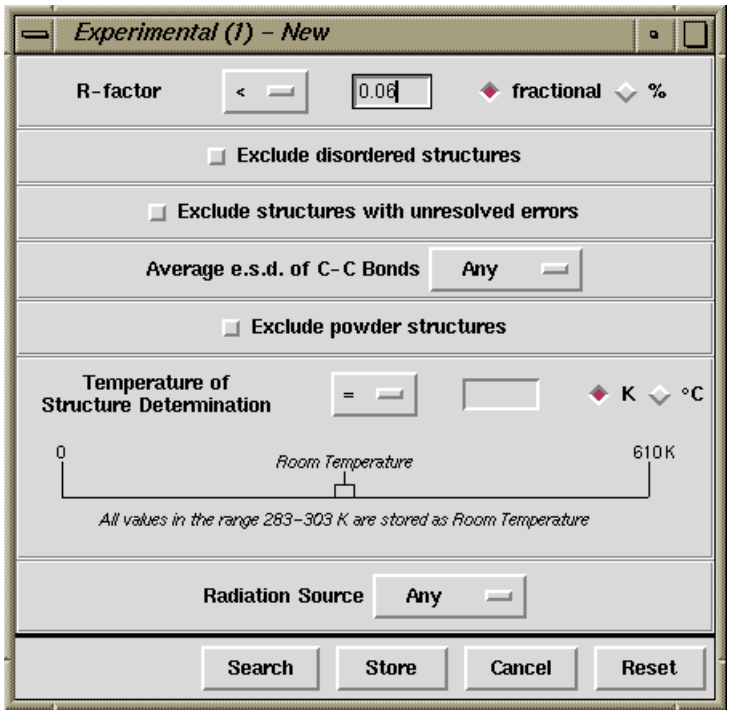


   - There are a number of alternative settings for this space group (e.g. P21/c and P21/a are the same space group but with a different choice of cell axes). We can choose to find all the alternative settings or just the P21/c setting. The former is almost always the sensible choice. So, ensure that the **Include alternative settings of selected space group** check-box is switched on.

- Hit **Store**.

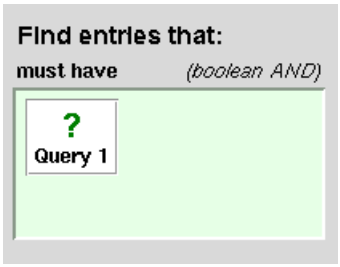
4. Set up a query for neutron diffraction structures.
   - Hit the **Experimental** button.
   - Change the *Radiation Source* from **Any** to **Neutron**.
   - Hit **Store**.

5. Set up a query for structures with R factors < 0.06.
   - Hit the **Experimental** button.
   - Change the button next to *R-factor* from = to <.
   - Type *0.06* into the input box, thus specifying structures with R factors below 6%:

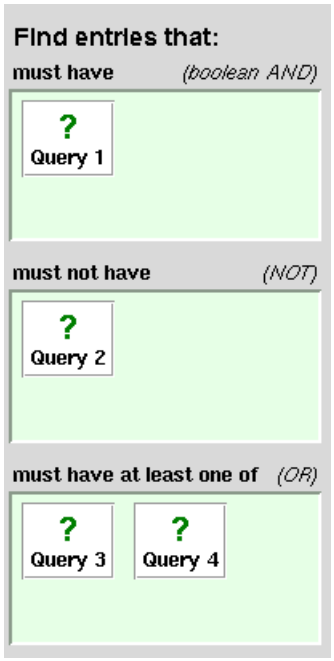


   - Hit **Store**.

6. Combine the queries.
   - Click on *Combine Queries*.
   - Position the cursor over the top query icon (i.e. the white square next to the formula query, labelled *Query 1*), press down on the left-hand mouse button, and drag the icon into the box labelled *must have*:



- Drag the space-group query icon (*Query 2*) into the box labelled *must not have*.
- Drag both of the remaining queries into the box labelled *must have at least one of*:



7. Run the search.
   - Hit the **Search** button at the bottom left-hand corner of the window.
   - The pop-up confirms that the search will be for structures that satisfy Query 1 (i.e. have Os1-2 O>6), do not satisfy Query 2 (i.e. do not belong to the space group P21/c) and satisfy Query 3 or Query 4 (i.e. determined by neutron diffraction or with an R-factor < 0.06; or both).
   - Hit **Start Search** to begin the search.

This ends the tutorial.

22 APPENDIX A: GLOSSARY AND TECHNICAL DETAILS

This section is a glossary of information fields and terminology you may come across when using ConQuest.

Absolute Configuration (see page 210)
Angles (see page 210)
Aromatic Bonds (see page 211)
Asymmetric Unit (see page 211)
Atomic Charge (see page 212)
Atomic Coordinates (see page 212)
Atomic Displacement Parameters (ADPs) (see page 213)
Atomic Labels (see page 213)
Authors' Names (see page 213)
Average e.s.d. of C-C Bonds (see page 214)
Azides (see page 214)
Bioactivity (see page 215)
Bond Type Conventions (see page 216)
Centroids (see page 217)
Chemical Diagram (see page 217)
Chemical Formula (see page 218)
CIF Format (see page 218)
CIFMIF Format (see page 219)
Comments (see page 219)
Compound Name Conventions (see page 219)
Compound Name Keywords Indicating Structure Type (see page 220)
COORD Format (see page 221)
Coordination Geometry (see page 221)
Crystal Chemical Unit (see page 221)
Crystal Systems (see page 222)
Cyclicity (see page 222)
Delocalised Double Bonds (see page 223)
Density (see page 223)
Deposition Numbers (see page 224)
Disordered Structures (see page 224)
Distances (see page 225)
Drug Activity (see page 225)
Dummy Points (see page 225)
Element Group Symbols (see page 225)
FDAT Format (see page 227)
Flack Parameter (see page 227)

Habit (see page 227)
Hydrogen Atoms in 2D/3D Substructure Searching (see page 227)
Hydrogen Atom Location in Crystal Structure Analyses (see page 228)
Inorganic Crystal Structure Database (ICSD) (see page 229)
Internal Comparison (see page 229)
Journal Name (see page 229)
Journal Page (see page 230)
Journal Volume (see page 230)
Lattice Type (see page 231)
Logical Operators (see page 231)
Mercury (see page 232)
Metal Carboxylates (see page 232)
Metal-Metal Bonds (see page 232)
Metals and Alloys Crystallographic Database (CRYSTMET) (see page 232)
Missing Atoms (see page 233)
Mol2 Format (see page 233)
Molecule (see page 233)
Neutron Diffraction (see page 233)
No Errors (see page 233)
No Ions (see page 234)
Nucleic Acids Database (see page 234)
Number of Atoms with 3D Coordinates (see page 234)
Number of Chemical Units (molecules, ions, etc.) (see page 234)
Occupancy Factor (see page 234)
Only Organics (see page 235)
Only Organometallics (see page 235)
pdb Format (see page 235)
Physical and Chemical Properties (see page 235)
Pi-Bonds (see page 236)
Planes (see page 236)
Pluto (see page 236)
Polymeric Bonds (see page 236)
Polymorphism (see page 237)
Powder Data (see page 237)
PreQuest (see page 237)
Private Communications (see page 237)
Protein Data Bank (see page 237)
Reduced Cell Parameters (see page 238)
Refcode (entry ID) (see page 238)
R-factor (see page 239)
Rogers Parameter (see page 239)

SD Format (see page 240)
Sohncke Groups (Chiral Space Groups) (see page 240)
Source of Crystal (see page 240)
Space Group (see page 240)
Temperature Factor (see page 241)
Temperature of Structure Determination (see page 241)
Torsion Angles (see page 242)
Unit Cell Parameters (see page 242)
Unresolved Errors (see page 243)
Van der Waals Radii (see page 243)
Variable Bond Type (see page 246)
Variable Element Type (see page 246)
Vectors (see page 246)
Vista (see page 246)
Z (Number of Molecules per Unit Cell) (see page 246)
Z' (Number of Molecules per Asymmetric Unit) (see page 247)

Absolute Configuration

The phrase *absolute configuration* is included in an entry if the absolute configuration of the structure has been determined by X-ray methods. This can only be done if the compound crystallises in one of the 65 Sohncke Groups (Chiral Space Groups) (see page 240).

To Search:
- Select the **All Text** button in the *Build Queries* pane (see GENERAL TEXT SEARCHES, page 135)

Related Topics
- Flack Parameter (see page 227)
- Internal Comparison (see page 229)
- Rogers Parameter (see page 239)

Angles

Angles (in degrees) can be calculated between atoms, dummy points, centroids, planes or vectors.

Related Topics
- Distances (see page 225)
- Torsion Angles (see page 242)

Aromatic Bonds

An *aromatic* bond type is used for the ring bonds of benzenoid systems, 6-membered aromatic heterocycles and cyclopentadienyl systems. However, 5-membered heterocycles such as thiophene, furan, etc. are normally coded with single and double bonds, as are annulenes.

A few hundred entries contain the 6-membered carbon ring with alternating single and double bonds. These structures are of 3 main types:

- Metal complexes where pi-bonding between the metal and the 6-membered carbon ring involves only 2 or 4 of the 6 carbon atoms.
- Phthalocyanines and other benzoporphines.
- Fullerenes - these structures always contain the root *fuller* in the compound name, so it is safer to search for them with a compound name search (see Using the Compound Name and Chemical Class Dialogue Box, page 116); it is also easier than trying to construct a fullerene diagram!

Related Topics

- Bond Type Conventions (see page 216)
- Chemical Diagram (see page 217)
- Delocalised Double Bonds (see page 223)
- Pi-Bonds (see page 236)
- Variable Bond Type (see page 246)

Asymmetric Unit

A crystal structure consists of a basic motif that is repeated in 3D space by the symmetry operators of the crystallographic space group. A crystallographer determines the coordinates of the atoms in this basic motif, called the *asymmetric unit*. It is the smallest part of a crystal structure from which the complete structure can be built using space group symmetry.

The asymmetric unit may consist of only one molecule or ion, part of a molecule, or several molecules that are not related by crystallographic symmetry. (This is why you sometimes see more than one molecule in the ConQuest 3D visualiser when you view results.) For example, consider structures of formula $C_{12}\ H_{18}\ N_4\ O_2$:

- If the asymmetric unit contains one molecule, the crystallographer must determine the coordinates of 36 atoms.
- If the asymmetric unit contains two molecules, the crystallographer must determine the coordinates of 72 atoms.

- If the asymmetric unit is half a molecule, this implies that the molecule possesses symmetry coincident with a crystallographic symmetry element. For example, the molecule might possess a mirror plane, so that half of the atoms are related to the other half by symmetry. In this case, the crystallographer must determine the coordinates of only 18 atoms.

  Related Topics
  - Atomic Labels (see page 213)
  - Crystal Chemical Unit (see page 221)
  - Molecule (see page 233)
  - Z (Number of Molecules per Unit Cell) (see page 246)
  - Z' (Number of Molecules per Asymmetric Unit) (see page 247)

Atomic Charge

Where appropriate, atoms are assigned formal charges. In the CSD there is no concept of delocalised charge - charges must be associated with specific atoms and must be integral. Formal charges are often rather arbitrary, particularly when assigned to metal atoms, so their use in substructure searching is not recommended except for simple organic structures.

Related Topics
- Chemical Diagram (see page 217)

Atomic Coordinates

Atomic coordinates are not available for about 10% of entries in the CSD. This situation usually arises for structures published as short communications in certain journals, or if data has not been received from an author.

Atomic coordinates may be written out as fractions of the unit cell edges (i.e. x/a, y/b, z/c.; so-called *fractional* coordinates) or as orthogonalised, Cartesian coordinates.

To Search *(to exclude entries which have no atomic coordinate data in the CSD*):
- Select the appropriate filter in the *Search Setup* dialogue box (see Applying Filters, page 143).

Related Topics
- CIF Format (see page 218)
- CIFMIF Format (see page 219)
- COORD Format (see page 221)
- FDAT Format (see page 227)
- Mol2 Format (see page 233)

- pdb Format (see page 235)
- SD Format (see page 240)

Atomic Displacement Parameters (ADPs)

Atomic displacement parameters are not currently available in the CSD.

Atomic Labels

Each atom in a structure has a label, which consists of the element symbol followed by a number and (sometimes) a prime. CSD atom labels are normally identical to, or closely related to, those used by the author.
Atoms generated by symmetry (i.e. not belonging to the asymmetric unit) will have an extra letter at the end of their label, which indicates which symmetry operation was used.

Related Topics
- Crystal Chemical Unit (see page 221)

Authors' Names

Typical author names are:

- A.B.Smith
- A.T.McPhail
- J.J.MacDougall
- C.J.O'Connor
- P.Murray-Rust
- F.de Bergevin
- J.-P.Dubois
- W.Van Riel
- D.van der Helm
- Yu.T.Struchkov

Authors' names are stored exactly as given in the paper, with forenames abbreviated and stored as initials, e.g. *F.Allen* and *F.H.Allen* may both occur. The use of initials in queries can therefore result in a loss of relevant hits.

Some author names may include dynastic tags, e.g. *S.S.Simons Junior, A.J.Arduengo III*.

Authors' names are stored in the CSD without umlauts, accents, etc. However, some journals express an umlaut by a following *e*; thus *Müller, Sänger*, etc., may sometimes appear as *Mueller, Saenger*, etc.

in the CSD.

Chinese, Korean and Malaysian names are usually stored in full, e.g. *Bing Bing Chang, Jung Mi Shin.* Occasionally, however, Oriental names are recorded with initials, e.g. *H.S.Kim.* These variations are due to different journal conventions.

Names from non-Roman alphabets are not always transliterated in the same way in different papers, e.g. *Belskii, Belsky, Belskij*; query strings should be restricted to the root of the name, *Belsk*.

There is some inconsistency in the database in the handling of 2-letter initials, e.g. *Yu.T.Struchkov* might occasionally be stored as *Y.T.Struchkov.*

To Search:

- Select the **Author/Journal** button in the *Build Queries* pane (see AUTHOR AND JOURNAL SEARCHES, page 113).

Average e.s.d. of C-C Bonds

The average estimated standard deviation (e.s.d.) [standard uncertainty (s.u.)] of the carbon-carbon bond lengths in a structure provides a rough measure of precision. In searching the CSD, you can either accept any value or restrict hits to values < 0.005Å. The latter option limits hits to more precise structure determinations.

In certain entries, if no carbon-carbon e.s.d.s are available, then the value may be derived from C-N, C-O, N-N, N-O, O-O e.s.d.s

To Search:

- Select the **Experimental** button in the *Build Queries* pane (see Using the Experimental Dialogue Box, page 133).

Azides

Azide bond types are inconsistent in the CSD; sometimes one canonical form is used, sometimes another. The safest procedure is to search for *X...N...N...N* where ... is a bond of type 1 or 2 or 3.

Related Topics

- Bond Type Conventions (see page 216)
- Chemical Diagram (see page 217)
- Variable Bond Type (see page 246)

Bioactivity

A number of keywords are stored in the *Bioactivity* field of the CSD. Examples of these can be found in the *All Text* scrolling list and some more common examples can be found in the table below:

adrenoreceptor	anticoagulant	antiinflammatory	cancer	inotropic
agonist	anticonvulsant	antileukaemi(a)(c) antileukemic	carcinogenic	insecticidal
analeptic	antidepress(ant)(ion)(ive)	antimalarial	cardiovascular	lymphocytic
analgesic	antidiabetic	antimetastatic	cholinergic	metalloproteinases
antagonist(ic)	antidopaminergic	antineoplastic	cytostatic	neurotoxicity
anthelmintic	antiestrogen	antiparasit(e)(ogenic)	cytotoxic(ity)	osteoporosis
antiAIDS	antifertility	antiparkinson	drug	psychostimulant
antiadrenergic	antifungal	antipsychotic	fungicidal	psychotropic
antiallergic	antigenic	antipyretic	herbicidal	receptor(s)
antiandrogen(ic)	antiglaucoma	antispasmodic	hypercholesterolemia	retroviral
antiarrhythmic	antigonadotropic	antitumour	hyperglycaemia	teratogenic
antiasthmatic	antihaemorrhagic	antiulcer	hypocholinergic	therapeutic
antibacterial	antihistamine	antiviral	hypoglycaemic	trypanocidal
antibiotic	antihyperglycemic	anxiolytic	hypolipidermic	vaso(co)constrict(ion)(ive)
anticancer	antihypertens(ion)(ive)	beta-adrenergic	hypotension	vasodialt(ive)(or)
anticholinergic	antiinfective	beta-blocking	immunosuppress(ant)(ive)	

Please note that both English and American spellings are used. It is advisable to search for both if you are unsure which should be used.

Related topics

- Drug Activity (see page 225)

Bond Type Conventions

It is necessary to understand the CSD bond-type conventions in order to do accurate substructure searching. For example, a benzene ring in the CSD is coded using an aromatic bond type rather than alternate single and double bonds. Thus, a substructure search for benzene will fail unless the substructure is drawn with aromatic bonds.

The CSD bond types are: single, double, triple, quadruple, aromatic, polymeric, delocalised double and pi.

The following conventions are used for common functional groups:

terminal carbonyl	−C≡O
cyano	−C≡N
cyanato	−O−C≡N
thiocyanato	−S−C≡N
isocyanato	−N=C=O
isothiocyanato	−N=C=S
nitro	O=N=O

Problems with substructure searching are very often due to bond-type issues, either a misunderstanding of CSD conventions or an inconsistency in the CSD itself. These inconsistencies principally apply to the structure types given below in related topics.

Related Topics

- Aromatic Bonds (see page 211)
- Azides (see page 214)
- Chemical Diagram (see page 217)
- Delocalised Double Bonds (see page 223)
- Metal Carboxylates (see page 232)
- Metal-Metal Bonds (see page 232)
- Pi-Bonds (see page 236)
- Polymeric Bonds (see page 236)
- Variable Bond Type (see page 246)

Centroids

Centroids can be calculated for sets of atoms.

Chemical Diagram

The chemical diagram is stored as a set of atom and bond properties. This representation forms the basis for 2D chemical substructure searches and, in combination with geometrical constraints, is also used in 3D searching.

Because the structures in the CSD cover a wide range of chemistry, a number of conventions have been employed in coding some of these properties, particularly chemical bond types. Information about these conventions and their implications for the coding of substructure queries can be found in the list of related topics.

Atom properties:

- Atomic Charge (see page 212)
- Cyclicity (see page 222)
- Element symbol
- Number of bonded atoms, excluding terminal hydrogens
- Number of terminal hydrogens

Bond properties:

- Bond type (see Bond Type Conventions, page 216)
- Cyclicity (see page 222)

To Search:

- Queries involving substructures are drawn in the *Draw* window (see Layout of the Draw Window, page 47)

Related Topics

- Aromatic Bonds (see page 211)
- Atomic Charge (see page 212)
- Azides (see page 214)
- Cyclicity (see page 222)
- Delocalised Double Bonds (see page 223)
- Element Group Symbols (see page 225)
- Hydrogen Atoms in 2D/3D Substructure Searching (see page 227)
- Metal Carboxylates (see page 232)
- Pi-Bonds (see page 236)

- Polymeric Bonds (see page 236)
- Variable Bond Type (see page 246)
- Variable Element Type (see page 246)

Chemical Formula

Since a crystal structure may contain more than one type of molecule, there are basically two sorts of formulae in the CSD:

- The formula of an individual molecule, e.g. $C_{10}\,H_{18}\,N_2\,Ni_1\,O_5$.
- The *sum formula*, i.e. the formula of all the different molecules in a structure added together. For example, the dihydrate $C_{10}\,H_{18}\,N_2\,Ni_1\,O_5,\,2(H_2O)$ would have a sum formula of $C_{10}\,H_{22}\,N_2\,Ni_1\,O_7$.

Both types of formula can be searched for in ConQuest.

The letter *n,* which may sometimes be seen in formulae displayed in *View Results*, indicates that the compound is polymeric, e.g. *(C3 H3 O6 Sc1)n*. The formula enclosed in the brackets is that of the monomer repeat unit.

To Search*:*

- Select the **Formula** button in the *Build Queries* pane (see Using the Formula Dialogue Box, page 123). (Exact searches involving charges and multipliers can be done using the **All Text** option (see Using the All Text Dialogue Box, page 135), but they run slowly.)

Related Topics

- Element Group Symbols (see page 225)

CIF Format

The Crystallographic Information File (CIF) format was developed as the standard crystallographic data exchange format (Hall, Allen and Brown, *Acta Cryst.*, **A47**, 655, 1991; *http://www.iucr.org/iucr-top/cif/home.html*).

Related Topics

- CIFMIF Format (see page 219)

CIFMIF Format

cifmif is an output format option for structure data from the CSD. This format combines CIF crystallographic data names with chemical data names from the Molecular Information File (MIF: Allen, Barnard, Cook and Hall, *J.Chem.Inf.Comput.Sci.*, **35**, 412, 1995).

Related Topics
- CIF Format (see page 218)

Comments

A great variety of comments are included in the CSD, e.g.

- There appears to be an error associated with the coordinates of C17.
- Solvate coordinates have not been reported.
- Dx is reported as 1.39; we calculate 1.50.
- Formula weight is reported in the paper as 552; we calculate 718.
- IUCr br1217
- IUC9900093

The last two are examples of Deposition Numbers (see page 224).

To Search:
- Select the **All Text** button in the *Build Queries* pane (see All Text Searching, page 135).

Compound Name Conventions

Compound names usually follow the rules of standard chemical nomenclature. Occasionally, a trivial, drug or trade name might be used.

A few conventions are used in constructing compound names:

- Lower-case Greek characters are spelt out, e.g. *alpha*. Upper-case Greek characters are spelt out and prefixed by *c*, e.g. *cdelta.*
- The names of elements Al, Cs and S are spelt *aluminium, cesium, sulfur.*
- Bridging ligands in polymeric metal coordination complexes are identified by the bridging indicator *mu*, with the polymer identified by the prefix *catena*, e.g. *catena-((mu$_2$-2,5-dihydroxy-p-benzoquinonato)-zinc).*
- Names of hydrates will contain the words *hemihydrate, monohydrate, dihydrate*, etc., otherwise, just *hydrate* if the multiplier is a non-integer value.
- If other solvents are present, the name will contain the word *solvate*; *clathrate* is used for

solvates which are clathrated, as in host-guest compounds.

- Deuterated species will always contain the name characters *deuter.*

To Search*:*

- Select the **Name/Class** button in the *Build Queries* pane (see Compound Name Searching, page 116).

Related Topics

- Compound Name Keywords Indicating Structure Type (see page 220)

Compound Name Keywords Indicating Structure Type

To assist searching, some types of compounds always contain certain words in their compound name:

- calixarene (see below)
- catenane
- clathrate
- cyclophane (see below)
- electride
- fullerene (see below)
- hydrate, monohydrate, dihydrate, etc.
- pseudopeptide or pseudo-peptide
- radical
- retropeptide or retro-peptide
- rotaxane
- trefoil knot

There are a number of variants of the word c*yclophane*, e.g. *cyclophanyl, cyclophanone,* etc. The safe procedure is to search for the root *cyclophan*. To find compounds such as cyclo-octaphane or cyclo-octacosaphane, it is best to search for the roots *cyclo-octa* and *phane*.

There are a number of variants of the word *calixarene*, e.g. *calix(4)crown, calix(4)resorcinarene* etc. The safe procedure is to search for the root *calix*.

There are a number of variants of the word *fullerene*, e.g. *fulleride, fullerenyl,* etc. The safe procedure is to search for the root *fuller.*

Related Topics

- Compound Name Conventions (see page 219)

COORD Format

The COORD format (file extension *.cor*) is a simple ASCII listing of 3D coordinates and, where appropriate, cell data. You may choose to write out orthogonal or fractional coordinates (see Atomic Coordinates, page 212), either for all the atoms in the Crystal Chemical Unit (see page 221) or just the atoms that matched the search substructure.

Coordination Geometry

Metal coordination geometries are not explicitly stored in the CSD. To search for a particular coordination geometry (e.g., square planar rather than tetrahedral), you must do a 3D substructure search (see Basic Guide to 3D Substructure Searching, page 79).

Crystal Chemical Unit

In some cases, molecular symmetry coincides with the symmetry elements of a crystallographic space group. For example, a molecule might have a mirror plane which coincides with a crystallographic mirror plane in the space group. In the crystal structure, the molecule has *exact m-symmetry*, the Asymmetric Unit (see page 211) is half a molecule, and the crystallographer only determines the coordinates of half of the atoms in the molecule.

In this situation, the atoms in the asymmetric unit plus the symmetry-generated atoms forming the other half of the molecule collectively form the *crystal chemical unit*.

If you request output of such a structure, the coordinates of the entire molecule (i.e. the crystal chemical unit) will be written out, not just the coordinates of the asymmetric unit. The symmetry-generated atoms can be distinguished from the atoms in the asymmetric unit because their atom labels will end in a letter. (The letter indicates which symmetry operation was applied to generate the atom.)

Related Topics

- Asymmetric Unit (see page 211)
- Atomic Labels (see page 213)
- Molecule (see page 233)
- Z (Number of Molecules per Unit Cell) (see page 246)
- Z’ (Number of Molecules per Asymmetric Unit) (see page 247)

Crystal Systems

Any crystal must belong to one of seven crystal systems, the symmetry of which introduces constraints on the possible values of the Unit Cell Parameters (see page 242). The seven crystal systems and the distribution of space groups in them are:

Crystal System	*Space Groups*
Triclinic/Anorthic	1-2
Monoclinic	3-15
Orthorhombic	16-74
Tetragonal	75-142
Trigonal/Hexagonal trigonal hexagonal	 143-167 with hexagonal axes 168-194
Rhombohedral trigonal	 146-167 with rhombohedral axes
Cubic	195-230

To Search*:*

- Select the **Space Group** button in the *Build Queries* pane (see Using the Space Group Dialogue Box, page 126).

Related Topics

- Space Group (see page 240)
- Unit Cell Parameters (see page 242)

Cyclicity

Cyclicity refers to whether or not an atom or bond is part of a ring, i.e. *cyclic* or *acyclic*.

Related Topics

- Chemical Diagram (see page 217)

Delocalised Double Bonds

Delocalised double bonds are used in the CSD to describe the bonding in some delocalised structures. An example is metal-coordinated carboxylate. Rather than drawing one of the C-O bonds as single and one double (i.e. O-C=O), both bonds are assigned the *delocalised double* bond type (O--C--O). This is an attempt to reflect the fact that the C-O bonds are equivalent to each other.

The correct bond types must be specified in substructure searching, otherwise structures will not be found. Unfortunately, the CSD is often inconsistent in its use of delocalised double bonds. For example, about 50% of metal-bound acetylacetonato ligands are coded with the 6-membered ring having 4 single bonds and 2 double bonds. The other 50% have 2 single bonds and 4 delocalised double bonds. Thus, the safest way to conduct a search for acetylacetonato complexes is to use a variable bond type when drawing the substructure, allowing the acetylacetonato bonds to be single, double or delocalised-double.

Related Topics

- Aromatic Bonds (see page 211)
- Azides (see page 214)
- Bond Type Conventions (see page 216)
- Chemical Diagram (see page 217)
- Metal Carboxylates (see page 232)
- Metal-Metal Bonds (see page 232)
- Pi-Bonds (see page 236)
- Polymeric Bonds (see page 236)
- Variable Bond Type (see page 246)

Density

This is the density of the crystal, calculated from the reported chemical formula and unit cell data, using the relationship:

Density = (1.66 x formula weight x Z) / unit cell volume

where Z is the number of molecules in the unit cell.

To Search*:*

- Select the **Z/Density** button in the *Build Queries* pane (see Using the Z/Density Dialogue Box, page 131).

Deposition Numbers

Many journals have an arrangement with the CCDC whereby the author deposits the crystallographic data at the CCDC in advance of publication and receives a 6-digit CCDC deposition number (e.g. CCDC 103296) which is printed in the paper. The number is stored in the corresponding CSD entry and is a useful search term for finding the entry.
In certain cases, journals print their own deposition codes, e.g.

- IUCR bk1432 (used for Acta Cryst., sections B and C).
- IUC9900104 (used for CIF-Access papers in Acta Cryst., Section C).
- CCDC 186/456 (used by the Royal Society of Chemistry; 186 is the CCDC journal identifier for J. Chem.Soc., Dalton Transactions).
- CSD 400922 (used in the data depository maintained by FIZ-Karlsruhe).

To Search*:*

- Select the **Author/Journal** button in the *Build Queries* pane (see CCDC number (deposition number), page 115).
- For IUCr deposition numbers, select the **All Text** button in the *Build Queries* pane (see Using the All Text Dialogue Box, page 135).

Disordered Structures

Disordered structures display a lack of regularity. For example, each of the F atoms in the trifluoromethyl group -CF3 might be randomly distributed between 2 sites. This means that the crystallographer will report two sets of coordinates for each F atom. In some cases, two alternative sites are occupied equally; in other cases, there is a major site and a minor site. Disorder can involve more than two sites and it can also involve a whole molecule. Disordered atoms can be hidden when viewing the structure in 3D (see Changing Display Styles and Display of Hydrogen Atoms, page 34).

The nature of the disorder is briefly described in a text comment, e.g. *The anion is disordered over two positions with equal occupancy.*

To Search *(to reject disordered structures):*

- Select the **Experimental** button in the *Build Queries* pane (see Using the Experimental Dialogue Box, page 133).
- Use the *Filters* menu in the *Search Setup* dialogue box (see Applying Filters, page 143).

Related Topics

- Occupancy Factor (see page 234)

Distances

Distances (in Å) can be calculated between atoms, dummy points, centroids and planes.

Related Topics
- Angles (see page 210)
- Torsion Angles (see page 242)

Drug Activity

Comments about drug activity or biological activity are included if the author indicates that the compound or a near-relative is a drug or has biological activity, e.g. *anticancer drug, derivative of cholinergic agent, antidepressant activity*. Searches can be restricted to entries with *Bioactivity* comments (see Specifying Required Fields, page 137).

Related Topics
- Bioactivity (see page 215)

To Search*:*
- Select the **All Text** button in the *Build Queries* pane (see All Text Searching, page 135).
- Select *Bioactivity* from the scrolling list in the **Required Fields** tab of the *All Text* dialogue box (see Specifying Required Fields, page 137).

Dummy Points

A dummy point is defined as the point lying on the extension of the vector from atom A to atom B, at a distance *d* Å from atom B.

Element Group Symbols

The following preset element group symbols can be used for specifying substructures containing variable atom types. For example, a search for the substructure *1B-Cl* would find structures containing any of the bonds Cu-Cl, Ag-Cl or Au-Cl. Element group symbols can also be used to generalise element and formula searches.

Element group symbol	*Elements contained in the group*
HD	H D
1A	Li Na K Rb Cs Fr
2A	Be Mg Ca Sr Ba Ra
3A	B Al Ga In Tl
4A	Si Ge Sn Pb
5A	N P As Sb Bi
6A	O S Se Te Po
7A	F Cl Br I At
8A	He Ne Ar Kr Xe Rn
1B	Cu Ag Au
2B	Zn Cd Hg
3B	Sc Y La Ac
4B	Ti Zr Hf
5B	V Nb Ta
6B	Cr Mo W
7B	Mn Tc Re
8B	Fe Co Ni Ru Rh Pd Os Ir Pt
8X	Fe Ru Os
8Y	Co Rh Ir
8Z	Ni Pd Pt
1M	1A + 2A
2M	Al Ga In Tl Ge Sn Pb Sb Bi Po
3M	1M + 2M
4M	1M + 2M + TR + LN + AN i.e. all metallic elements
NM	non-metallic elements
Any	all elements
R	any alkyl group i.e. C or H

TR	all transition metals, but not the lanthanides or actinides
1R	the first row transition metals i.e. Sc to Zn
2R	the second row transition metals i.e. Y to Cd
3R	the third row transition metals i.e. La to Hg, not the lanthanides
L	the lanthanides i.e. Ce to Lu, but not including La
AN	the actinides i.e. Th to Lr, but not including Ac

FDAT Format

FDAT is a file format used for storing the crystallographic data contained in an entry. Its use is no longer recommended but, as many people have written programs that read the format, it is still available as an output option.

Flack Parameter

This is a parameter (Flack, *Acta Cryst*., **A39**, 879, 1983) that can be used in the determination of Absolute Configuration (see page 210).

Related Topics
- Absolute Configuration (see page 210)

Habit

Crystal habit (i.e. the shape of the crystal, such as *needle* or *plate*) is stored for some entries.
The most common words or phrases used to describe habit are included in the scrolling list in the **All Text** dialogue box (see All Text Searching, page 135).

To Search*:*
- Select the **All Text** button in the *Build Queries* pane, click on the **Required Fields** tab and select *Habit* from the scrolling list (see Specifying Required Fields, page 137). Note, this will return hits only for entries where habit information is available.

Hydrogen Atoms in 2D/3D Substructure Searching

Hydrogen atoms can be difficult to locate by X-ray crystallography and hence their coordinates are sometimes missing or incomplete in CSD entries. However, they are always completely defined in the Chemical Diagram (see page 217) that forms the basis for 2D chemical substructure searching, and these searches will proceed accurately under all circumstances, *so long as no 3D geometrical parameters are defined.* However, if a moiety in a 2D diagram displays an agostic or bridging M-H-CH_2-R interaction, the agostic H atom must be defined explicitly to retrieve relevant entries from the CSD (see Automatic Addition of Hydrogen Atoms, page 58) and (see Adding Hydrogen Atoms

Manually, page 59).

If H-atoms form part of a substructure and any sort of 3D parameter (atom labels, distances, angles, etc.) is defined, then 3D coordinates must be accessed, and the missing or incomplete set of H-coordinates *can cause you to miss hits*. For example, if you are looking for structures with S-H bonds in the range 1.0-1.3 Å, you will not locate any structures in which the H atom position is undefined, because it is impossible to compute the S-H distance in these cases.

Whether structures are missed may depend on how the hydrogens are defined when drawing the substructure. Hydrogen atoms can be drawn in two ways:

- *Implicitly,* e.g. by right-clicking on a non-hydrogen atom and selecting the menu item **Hydrogens** from the pull-down menu; this will be shown as, e.g. SH_1. Hydrogens added in this way are terminal atoms i.e. they cannot be used to retrieve hits with bridging H atoms.
- *Explicitly,* by actually drawing them; this will be shown as, e.g. *S-H*. Hydrogens added in this way can be used to define hydrogen bridging positions.

If a 3D parameter has also been defined, then these two alternative ways of defining hydrogen atoms will give different results, *even if that 3D parameter does not involve the hydrogen atom*.

Thus:

- If the hydrogen has been defined implicitly, it does not matter whether its coordinates are present or absent.
- If the hydrogen atom has been drawn explicitly, its coordinates must be present in the CSD, otherwise a structure will not be counted as a hit even if it contains the substructure you have drawn. In addition, H atoms are considered to form part of a substructure if they might be present by virtue of an *Any* atom or a multipick selection: even if you have not defined H atoms explicitly you may obtain unexpected results.

Hydrogen Atom Location in Crystal Structure Analyses

Hydrogen atoms can be difficult to locate using X-ray crystallography, especially if atoms of high atomic weights occur in the structure, and coordinates for some or all H-atoms may be absent in reported structures. In modern experiments, it is usual to either:

- model H-atoms into the structure at positions calculated using chemical knowledge (e.g. carbon hybridisation states), or
- to locate them in (difference) electron-density maps in the final stages of structure determination. A variety of H-atom refinement strategies are then employed.

Because X-rays *see* electron density, and the density around an H-atom is asphericallly displaced towards the atom to which it is bonded, X-H distances reported from X-ray analyses are usually significantly shorter than the true X-H internuclear distance (see, e.g. *Acta Cryst*., **B42**, 515, 1986).

However, hydrogen atoms are precisely located using neutron diffraction and, in generating geometrical descriptions of substructures involving H-atoms (particularly the geometry of hydrogen bonds), it is common to *neutron-normalise* the H-atom positions.

It is assumed that the experimentally determined X-H vector direction is correct, and the H-atom is then moved along this vector so that the X-H internuclear distance corresponds to the value determined by neutron diffraction. ConQuest will carry out this normalisation using either:

- C-H = 1.083Å, N-H = 1.009Å, and O-H = 0.983Å; these distances are averages (*J.Chem.Soc.Perkin Trans. 2*, S1-S19, 1987) taken over good neutron data available at that time, or
- X-H distances specified by the user.

Inorganic Crystal Structure Database (ICSD)

This is the inorganic counterpart of the Cambridge Structural Database (Bergerhoff, Hundt, Sievers and Brown, *J.Chem.Inf.Comput.Sci*., **23,** 66, 1983). The ICSD is produced by FIZ Karlsruhe, D-7514 Eggenstein-Leopoldshafen, Germany, in conjunction with the National Institute of Standards and Technology, U.S.A.

Internal Comparison

This is a method of determining Absolute Configuration (see page 210) by preparing a derivative in which a carbon atom of known chirality is introduced into the molecule during synthesis of the derivative. For example, the absolute stereochemistry of (+)-3-camphor is well known, so the formation of the (+)-3-camphor sulfonate derivative yields a new compound in which carbon centres of known chirality are present. The technique is rarely used.

Related Topics
- Absolute Configuration (see page 210)

Journal Name

Journal names in the CSD are normally the abbreviations adopted by the International Standards Organisation.

Some journals have changed their names over time, often by dividing into two or more sections; significant examples are listed in APPENDIX B: CHANGES IN JOURNAL NAMES (see page 248).

To Search:

- Select the **Author/Journal** button in the *Build Queries* pane (see Journal Name, page 114).

Related Topics

- Journal Page (see page 230)
- Journal Volume (see page 230)

Journal Page

The page number is the starting page of the publication. Most journal pages are numeric but sometimes they contain letters, e.g. *L25, 123S*.

For about 20 journals, mostly Russian and Chinese, each issue starts at page 1. To provide an unambiguous reference, we store the issue number as well as the page number in the journal page field, separating them by a hyphen, e.g. *89-3* indicates page 89 of issue 3.

Since 1997, *Acta Cryst.,Sect.C (Cr.Str.Comm.)* has reported so-called *CIF-access papers*. These are represented in the Table of Contents of each issue by a brief abstract and diagram. The actual paper does not appear in the printed issue but can be retrieved from the IUCr archive using the printed journal data validation number. An example of such a number is IUC9900004. In this case the journal page in the CSD would contain 9900004.

J.Chem.Res. is published in two parts - the synopsis (S) and the miniprint (M). Since this journal has no volume number, the page number for part S is stored in the journal volume field and the page number for part M is stored in the journal page field.

To Search:

- Select the **Author/Journal** button in the *Build Queries* pane (see Journal Page, page 115).

Related Topics

- Journal Name (see page 229)
- Journal Volume (see page 230)

Journal Volume

The journal volume is normally numeric but may occasionally contain letters, e.g. *C471*. However, in other cases the alphabetic part of a volume number is absorbed into the journal name, e.g. *Acta Cryst., Section B, volume 47*.

For most journals, the volume takes a single value for each journal year. However, some journals

have more than one volume per year and others have volume numbers that span a year change.

Some journals have no volume number, e.g. journals of the UK Royal Society of Chemistry, such as *J. Chem.Soc., Dalton Transactions*. Volume numbers are absent from PhD theses and private communications.

To Search*:*

- Select the **Author/Journal** button in the *Build Queries* pane (see Journal Volume, page 114).

Related Topics

- Journal Name (see page 229)
- Journal Page (see page 230)

Lattice Type

The lattice type refers to the choice of unit cell that has been made. If the smallest possible unit cell of a crystal has been chosen, the lattice type is called *primitive* (*P*). In some space groups, it is conventional to use a unit cell which is twice or four times the volume of the primitive cell. These non-primitive, or *centred*, lattice types may be *A, B or C-centred* (*A, B, C*), *body centred* (*I*) or *face-centred* (*F*). The first character of a space group symbol always defines the lattice type, e.g. *P21/c, Fdd2*.

It is essential to specify the lattice type correctly when doing a reduced cell-parameter search.

Related Topics

- Reduced Cell Parameters (see page 238)
- Space Group (see page 240)

Logical Operators

Three types of logical operators can be used to combine CSD queries: Boolean *AND*, *OR* and *NOT*. For example, a search can be done for structures:

- containing selenium *AND* with a density > 1.5; but
- *NOT* containing iron; and
- solved by neutron diffraction *OR* at a temperature below 200K.

This type of search is done in the *Combine Queries* part of the interface (see Complex Query Combinations, page 21).

Mercury

Mercury is a CCDC program for crystal structure visualisation which allows users to view either their own structures or those retrieved via a ConQuest search. Mercury is also available as a free download from: *http://www.ccdc.cam.ac.uk/products/csd_system/mercury/*

Metal Carboxylates

When both oxygen atoms of a carboxylate group are bonded to one or two metal atoms then, by convention, both C...O bonds are of type *delocalised double*. However, for simple acid salts, e.g. sodium acetate, our convention is to use localised single and double bonds.

The same principles apply to the analogous thiocarboxylates, dithiocarbamates, xanthate, thioxanthates etc.

Related Topics
- Bond Type Conventions (see page 216)
- Chemical Diagram (see page 217)
- Delocalised Double Bonds (see page 223)
- Variable Bond Type (see page 246)

Metal-Metal Bonds

There are two problems associated with metal-metal bonds:

- In some cases it is difficult to judge whether or not a bond exists.
- The bond order is often debatable.

Probably the safest approach is to search for structures with no metal-metal bond and also structures with a metal-metal bond of type 1 or 2 or 3 or 4, using a variable bond type.

Related Topics
- Bond Type Conventions (see page 216)
- Chemical Diagram (see page 217)
- Variable Bond Type (see page 246)

Metals and Alloys Crystallographic Database (CRYSTMET)

CRYSTMET contains crystallographic information relating to metal and alloy structures. Originally developed at the National Research Council in Ottawa, Canada, it is now maintained by Toth Information Systems, 2045 Quincy Avenue, Gloucester, Ontario K1J 6B2, Canada.

Missing Atoms

Atoms may be missing from CSD structures for several reasons:

- The crystallographer did not report them in the paper, i.e. only space group and cell parameters were given.
- Hydrogen coordinates are occasionally not reported in the paper or deposited at the CCDC with the rest of the atomic coordinates.
- Solvate atoms are occasionally not reported; a comment is given in this case.
- The crystallographer did not locate them, e.g in the case of a bridging hydrogen atom.
- CCDC staff may *suppress* one or more atoms. A common example is when the author reports two sets of atomic coordinates, for major and minor sites, in a disordered structure. In most cases the atoms of the minor site will be suppressed.

Related Topics

- Disordered Structures (see page 224)

Mol2 Format

mol2 is a molecular file format produced by Tripos Inc., 1699 South Hanley Road, St Louis, MO, 63144-2913 USA, and is a ConQuest output option.

Molecule

The word *molecule* in this document is used generically to refer to a molecule or ion. For example, sodium acetate monohydrate has 3 molecules: $C_2H_3O_2^-$, Na^+, H_2O.

Neutron Diffraction

A relatively small proportion of structures in the CSD have been determined by neutron rather than X-ray diffraction. The advantage of neutron diffraction is that it locates hydrogen atoms with good precision.

To Search:

- Select the **Experimental** button in the *Build Queries* pane (see Using the Experimental Dialogue Box, page 133).

No Errors

Selecting this filter restricts the search to entries which have atomic coordinates and do not contain unresolved errors.

To Search*:*

- Select the appropriate filter in the *Search Setup* dialogue box (see Applying Filters, page 143).

No Ions

Selecting this filter excludes structures that contain ions. Zwitterions will not be excluded.

To Search:

- Select the appropriate filter in the *Search Setup* dialogue box (see Applying Filters, page 143).

Nucleic Acids Database

The Nucleic Acids Database (Berman et al. *Biophys J*, **63,** 751, 1992; *http://ndb-mirror-2.rutgers.edu/*) contains structures of oligonucleotides and nucleic acids.

Number of Atoms with 3D Coordinates

The *number of atoms with 3D coordinates* is, more precisely, the number of atoms in the Crystal Chemical Unit (see page 221) with 3D coordinates. In other words, it is the number of experimentally-located crystallographically-independent atoms in the Asymmetric Unit (see page 211) plus any atoms that are symmetry-related and bonded to these.

To Search*:*

- Select the **Z/Density** button in the *Build Queries* pane (see Using the Z/Density Dialogue Box, page 131).

Number of Chemical Units (molecules, ions, etc.)

The *number of chemical units* refers to the number of molecules or ions etc. which are present in the crystal structure, e.g. selecting a value of 1 would retrieve all entries containing just one chemical unit.

To Search*:*

- Select the **Z/Density** button in the *Build Queries* pane (see Using the Z/Density Dialogue Box, page 131).

Occupancy Factor

A parameter that defines the partial occupancy of a given site by a particular atom. It is most frequently used to describe disorder in a portion of a molecule.

Related Topics

- Disordered Structures (see page 224)

Only Organics

Selecting this filter eliminates from the search any structure that contains a transition metal, lanthanide, actinide or any of Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po.

To Search*:*

- Select the appropriate filter in the *Search Setup* dialogue box (see Applying Filters, page 143).

Only Organometallics

Selecting this filter restricts the search to structures containing at least one transition metal, lanthanide, actinide, or any of Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po.

To Search*:*

- Select the appropriate filter in the *Search Setup* dialogue box (see Applying Filters, page 143).

pdb Format

pdb is the Protein Data Bank format for 3D structures and is a ConQuest output option.

Related Topics

- Protein Data Bank (see page 237)

Physical and Chemical Properties

Physical and chemical properties are sometimes recorded., e.g. *hygroscopic, air-sensitive, piezoelectric, phase transition at 123K, isomorphous with the Co analogue.* For a more complete list of possible search terms, see the scrolling list in the *All Text* dialogue box, or the *Required Fields* scrolling list, also available from the *All Text* dialogue box.

To Search*:*

- Select the **All Text** button in the *Build Queries* pane (see All Text Searching, page 135).
- Select the **All Text** button in the *Build Queries* pane, then click on the **Required Fields** tab and select the relevant field from the scrolling list (see Specifying Required Fields, page 137).

Pi-Bonds

Pi-bonds in the CSD refer to bonds between a metal and a ligand. For a cyclopentadienyl ring with no substituents or just acyclic substituents, a pi-bond is recorded between the metal atom and each of the ring C atoms. Similarly, butadiene would have 4 pi-bonds recorded to the metal atom.

Related Topics
- Aromatic Bonds (see page 211)
- Bond Type Conventions (see page 216)
- Chemical Diagram (see page 217)
- Delocalised Double Bonds (see page 223)
- Polymeric Bonds (see page 236)

Planes

Planes can be calculated through sets of atoms by least-squares fitting.

Pluto

A CCDC program for visualising crystal and molecular structures, hydrogen-bonding networks, etc. (Only available for Unix/Linux). Pluto is available as a free download from: *http://www.ccdc.cam.ac.uk/free_services/rpluto/*

Polymeric Bonds

Two kinds of polymers are found in the CSD:

- Coordination polymers, in which bridging between metals causes an indefinite extension of the structure.
- Organic polymers, i.e. molecules identified by the multiple repetition of one of more types of groups or atoms linked to each other, e.g. poly(vinyl chloride).

Both may be regarded as characterised by a repeating unit. Bonds between repeating units are assigned the *polymeric bond type*.

To Search *(to exclude structures containing polymeric bonds):*
- Select the appropriate filter in the *Search Setup* dialogue box (see Applying Filters, page 143).

Related Topics
- Chemical Diagram (see page 217)
- Bond Type Conventions (see page 216)
- Delocalised Double Bonds (see page 223)

- Metal Carboxylates (see page 232)

Polymorphism

Polymorphism is the occurrence of two or more crystalline forms of the same substance. CSD entries known to be polymorphic contain comments which include the word *polymorph*, e.g. *non-triboluminescent polymorph.*

To Search*:*

- Select the **All Text** button in the *Build Queries* pane, then click on the **Required Fields** tab and select the keyword *Polymorph* from the scrolling list (see Specifying Required Fields, page 137). Note that this will retrieve hits only where polymorph information is available.

Powder Data

A comment is included if a structure has been determined by powder diffraction, e.g *powder data*. This information will be included in the *Extra Information* field under the **Experimental** tab in the *View Results* window of ConQuest.

To Search*:*

- Select the **All Text** button in the *Build Queries* pane (see All Text Searching, page 135).

PreQuest

A CCDC program which allows users to convert their own crystal structures to CSD format, and hence create a database of in-house structures. This database can then be viewed and searched with ConQuest.

Private Communications

The CSD contains many private communications. In such entries, the journal name is *Private Communication.* There will be an indication of the source of the structure, e.g. C*ontribution from Department of Chemistry, University of Delaware, Newark DE 19716, USA*.

To Search:

- Select the **Author/Journal** button in the *Build Queries* pane (see Journal Name, page 114).

Protein Data Bank

The Protein Data Bank is maintained by the Research Collaboratory for Structural Bioinformatics (*http://www.rcsb.org/*) and is the single international repository for the processing and distribution of 3D macromolecular structure data, primarily determined experimentally by X-ray crystallography and NMR.

Reduced Cell Parameters

Since the choice of unit cell axes is often arbitrary, it is clearly desirable to conduct unit cell searches against cells which have been converted to a standard reduced form. This is *reduced cell searching.* The cell reduction process involves determining the primitive cell which has the shortest possible cell lengths - the *Buerger cell* (Mighell and Rodgers, *Acta Cryst*., **A36**, 321, 1980).

Example:

Unit cell:	12.132	14.662	13.153	90	127.84	90
Reduced cell:	11.154	12.132	14.662	90	90	111.36

In a number of cases, further restrictive rules have to be applied for a unique choice - the *Niggli cell* (P.Niggli *Handbuch der Experimentalphysik*, Vol.7, Part 1. Leipzig: Akademische Verlagsgesellschaft, 1928).

Unfortunately, although the Niggli cell is unique it can, under special circumstances, be unstable. A small variation in an input cell dimension may lead to a discontinuous change in the Niggli cell and this will give rise to problems for searches which allow for reasonable error bounds on cell dimensions. Fortunately only the cell angles are subject to abrupt change. This has lead to the suggestion (Andrews, Bernstein and Pelletier, *Acta Cryst*. **A36**, 248, 1980) that cell lengths alone be used for comparison purposes. This is what is done in ConQuest. Hence, two reduced cells will be held to match if their cell lengths are identical (within the specified tolerance), even if their cell angles are very different.

To Search:

- Select the **Unit Cell** button in the *Build Queries* pane (see Using the Unit Cell Dialogue Box, page 129).

Related Topics

- Unit Cell Parameters (see page 242)

Refcode (entry ID)

Each CSD entry has a unique identifier known as a *refcode*. The refcode contains six letters, e.g. ABACOF. If more than one study of a compound is present in the CSD, the second and subsequent entries will have two numbers after the six letters, e.g. ABACOF01, ABACOF10, ABACOF03.

Deuterated forms of a compound have the same six-letter code as the non-deuterated form.

Stereoisomers have different 6-letter codes.

To Search:

- Select the **Refcode (entry ID)** button in the *Build Queries* pane (see REFCODE SEARCHES, page 140).
- Select **View Databases** from the top level menu (see REFCODE SEARCHES, page 140).

R-factor

The crystallographic R-factor is the traditional figure of merit for crystal structures and provides a measure of how well the refined structure agrees with the experimental model. Authors often also report a weighted R-factor, wR. The value stored in the CSD is the lower of R and wR.

A rough guide to the quality of structure determinations is:

R-factor	*Quality*
0.01 - 0.03	Exceptional
0.03 - 0.04	Very high
0.04 - 0.05	High
0.05 - 0.07	Good
0.07 - 0.09	Average
0.09 - 0.10	Fair
0.10 - 0.15	Poor
> 0.15	Bad

To Search:

- Select the **Experimental** button in the *Build Queries* pane (see Using the Experimental Dialogue Box, page 133).
- Select the appropriate **Filter** in the *Search Setup* dialogue box (see Applying Filters, page 143).

Rogers Parameter

This is a parameter (eta: Rogers, *Acta Cryst*., **A37**, 734, 1981) that can be used in the determination of Absolute Configuration (see page 210).

Related Topics

- Absolute Configuration (see page 210)

SD Format

SD is a molecular file format produced by MDL Information Systems Inc., 14600 Catalina Street, San Leandro, CA 94577, USA, and is a ConQuest output option.

Sohncke Groups (Chiral Space Groups)

These 65 space groups contain only rotation or screw axes, and are the only space groups in which chiral structures can crystallise. They are space group numbers: 1, 3-5, 16-24, 75-80, 89-98, 143-146, 149-155, 168-173, 177-182, 195-199 and 207-214.

Related Topics
- Space Group (see page 240)
- Absolute Configuration (see page 210)

Source of Crystal

Comments are sometimes included to indicate how crystals were prepared e.g. *from the melt, dichloromethane/hexane*; if there are comments, these can be found in the *Additional Information* section under the **Crystal** tab of the *View Results* window. Comments to indicate the source of natural products are also occasionally given and can be found in the *Source* field under the **Chemical** tab in the *View Results* window, e.g. *Leaves of flemingia fruticulose wall (Leguminos ae).*

To Search:
- Select the All Text button in the Build Queries pane, then click on the Required Fields tab and select the keyword *Natural Source* from the scrolling list (see Specifying Required Fields, page 137).

Space Group

There are 230 possible arrangements of symmetry elements in the solid state. They are called *space groups* (see *International Tables for Crystallography,* Volume A, Kluwer Academic Publishers, 1983). Any crystal must belong to one (and only one) space group.

The space groups are numbered from 1 to 230 and each is represented by a space group symbol, e.g. space group number 19 has the symbol P212121.

For certain space groups, it is possible to choose the unit cell axes and/or origin in more than one way (*alternative settings*). This means that a given space group number can correspond to several space group symbols. For example, space group number 25 corresponds to Pmm2, P2mm and Pm2m.

The following trigonal space groups can be described with respect to hexagonal or rhombohedral axes:

146	R3
148	R-3
155	R32
160	R3m
161	R3c
166	R-3m
167	R-3c

In the CSD, the rhombohedral setting is identified by appending the letter *r* to the space group symbol, e.g. R-3cr.

In certain situations there is an unresolved ambiguity in the space group determination. For example, if the crystallographer cannot decide between Pmaa and P2aa then the space group is represented by the so-called aspect symbol, in this case P*aa. There are 127 aspect symbols having *CSD space group numbers* in the range 502-765.

To Search:

- Select the **Space Group** button in the *Build Queries* pane (see Using the Space Group Dialogue Box, page 126).

Related Topics

- Crystal Systems (see page 222)
- Unit Cell Parameters (see page 242)

Temperature Factor

Temperature factors are not currently available in the CSD.

Temperature of Structure Determination

A comment is included in the *Temperature* section under the **Experimental** tab in the *View Results* window if a structure has been determined at a temperature other than room temperature. Any structure determined in the range 283-303K is considered to be a room-temperature structure.

To Search:

- Select the **Experimental** button in the *Build Queries* pane (see Using the Experimental

Dialogue Box, page 133).

Torsion Angles

Torsion angles are used to describe conformations around rotatable bonds. The torsion angle between 4 atoms A-B-C-D is the angle by which the vector A-B must be rotated in order to eclipse the vector C-D when viewed along the vector B-C. Crystallographers usually express torsion angles in the range -180 to +180 degrees. According to convention (Klyne and Prelog, *Experientia* **16**, 521, 1960) a clockwise rotation is positive and an anti-clockwise rotation is negative.

Torsion angles may be calculated for any set of four objects, including atoms, centroids and dummy points.

Related Topics
- Angles (see page 210)
- Distances (see page 225)

Unit Cell Parameters

The unit cell is the basic building block of a crystal, repeated infinitely in three dimensions. It is characterised by:

- three vectors (*a, b, c*) that form the edges of a parallelepiped;
- the angles between the vectors (*alpha*, the angle between *b* and *c*; *beta*, the angle between *a* and *c*; *gamma*, the angle between *a* and *b*).

Depending on the crystal system (see Crystal Systems, page 222), there are sometimes restrictions on the values that unit cell parameters can take:

Triclinic/Anorthic	no restrictions	
Monoclinic		alpha = gamma = 90
Orthorhombic		alpha = beta = gamma = 90
Tetragonal	a = b	alpha = beta = gamma = 90
Trigonal/Hexagonal	a = b	alpha = beta = 90, gamma = 120
Rhombohedral	a = b = c	alpha = beta = gamma
Cubic	a = b = c	alpha = beta = gamma = 90

In the above table, alpha=gamma=90 for monoclinic corresponds to the *b-axis unique* setting. Two other settings are possible:

a-axis unique: beta = gamma = 90
c-axis unique: alpha = beta = 90

To Search*:*

- Select the **Unit Cell** button in the *Build Queries* pane (see Using the Unit Cell Dialogue Box, page 129).

Related Topics

- Crystal Systems (see page 222)
- Reduced Cell Parameters (see page 238)
- Space Group (see page 240)

Unresolved Errors

When a new entry is added to the CSD, checks are carried out to test for consistency between atomic coordinates and bond lengths. If errors are detected which cannot be fixed, the problem is sent to the author. In the meantime, the entry is archived to the CSD with a comment summarising the suspected problem.

Inevitably, some authors do not respond so the CSD contains entries with unresolved errors.

To Search *(to exclude entries with unresolved errors):*

- Select the **Experimental** button in the *Build Queries* pane (see Using the Experimental Dialogue Box, page 133).

Van der Waals Radii

Standard van der Waals radii are assigned to the common elements. They are taken from Bondi, *J.Phys.Chem*., **68,** 441, 1964. Other elements are assigned van der Waals radii of 2.0Å.

Standard van der Waals Radii (Å)

Element	***Radius***
Ag	1.72
Ar	1.88
As	1.85
Au	1.66
Br	1.85

Element	*Radius*
C	1.70
Cd	1.58
Cl	1.75
Cu	1.40
F	1.47
Ga	1.87
H	1.20
He	1.40
Hg	1.55
I	1.98
In	1,93
K	2.75
Kr	2.02
Li	1.82
Mg	1.73
N	1.55
Na	2.27
Ne	1.54
Ni	1.63
O	1.52
P	1.80
Pb	2.02
Pd	1.63
Pt	1.72
S	1.80
Se	1.90
Si	2.10
Sn	2.17

Element	***Radius***
Te	2.06
Tl	1.96
U	1.86
Xe	2.16
Zn	1.39

Variable Bond Type

ConQuest allows the user to specify that a bond in a substructure may be of variable type, e.g. *single, double or aromatic*. This is particularly useful if you are unsure about the bond-type conventions that are used in the CSD (e.g. the difference between *double* and *delocalised double* bonds), or if a substructure contains a group that has several possible valence-bond representations.

Related Topics

- Aromatic Bonds (see page 211)
- Bond Type Conventions (see page 216)
- Chemical Diagram (see page 217)
- Delocalised Double Bonds (see page 223)
- Metal Carboxylates (see page 232)
- Metal-Metal Bonds (see page 232)
- Pi-Bonds (see page 236)

Variable Element Type

ConQuest allows the user to specify that an atom can be a member of a preset element group, for example the element group symbol *6A* = O, S, Se, Te, Po (see Element Group Symbols, page 225). Alternatively the user can define a variable element type in the *Draw* window, e.g. element symbol *QA* = O or S (see Setting Variable Element Types, page 56).

Thus, if a particular search yields a low number of hits it might be appropriate to consider running the search again using variable element types. This can be very useful for the groups of transition elements.

Vectors

Vectors can be calculated between pairs of atoms.

Vista

A CCDC program for performing statistical analyses on ConQuest search results, e.g. calculation of averages, plotting of histograms and scattergrams.

Z (Number of Molecules per Unit Cell)

Although *Z* is normally referred to as the number of *molecules* in the crystallographic unit cell, the word *molecule* should more strictly be *formula unit*, e.g. would include both the metal-containing moiety and the water molecule in a compound of formula $C_{10}H_{10}Fe_1N_1O_2, H_2O$.

To Search:

- Select the **Z/Density** button in the *Build Queries* pane (see Using the Z/Density Dialogue Box, page 131).

Related Topics
- Z' (Number of Molecules per Asymmetric Unit) (see page 247).

Z' (Number of Molecules per Asymmetric Unit)

This is best explained with an example. Consider a simple molecule $C_{12}H_{18}N_4O_2$:

- If Z' = 1, the asymmetric unit contains 1 molecule and the crystallographer must determine the coordinates of 36 atoms.
- If Z' = 2, the asymmetric unit contains 2 molecules and the crystallographer must determine the coordinates of 72 atoms. The two molecules are *crystallographically independent*, i.e. they are not related by symmetry so do not have to have exactly the same geometry.
- If Z' = 0.5, the asymmetric unit is half a molecule and this implies that the molecule possesses symmetry coincident with a crystallographic symmetry element. The molecule might possess a mirror plane, so that one half of the atoms are related to the other half across the crystallographic mirror plane. In this case the crystallographer must determine the coordinates of only 18 atoms.

In crystal structures of solvates, ion pairs, clathrates and host-guest complexes, the word *molecule* should be taken to include the entire formula unit, e.g. both the metal-containing moiety and the water molecule in a compound of formula $C_{10}H_{10}Fe_1N_1O_2, H_2O$.

To Search:
- Select the **Z/Density** button in the *Build Queries* pane (see Using the Z/Density Dialogue Box, page 131).

Related Topics
- Z (Number of Molecules per Unit Cell) (see page 246).

23 APPENDIX B: CHANGES IN JOURNAL NAMES

Some journals have been split into separate parts at a certain date, or, in other cases, multiple parts of a journal have been merged together. This means that a particular journal name may only be valid for a certain period of time. The most important of these in the CSD are:

Acta Crystallogr.	1948-1967
Acta Crystallogr.,Sect.A(Cryst.Phys.Diffr.,Theor.Crystallogr.)	1968-1982
Acta Crystallogr.,Sect.B(Struct.Crystallogr.Cryst.Chem	1968-1982
Acta Crystallogr.,Sect.A (Found.Crystallogr.)	1983-
Acta Crystallogr.,Sect.B (Struct.Sci.)	1983-
Acta Crystallogr.,Sect.C (Cryst.Struct.Comm.)	1983-
Acta Crystallogr.,Sect.D (Biol.Crystallogr.)	1993-
Acta Crystallogr.,Sect.E (Struct.Rep.Online)	2001-
J.Chem.Soc.	1849-1877
	1926-1965
J.Chem.Soc.A	1966-1971
J.Chem.Soc.B	1966-1971
J.Chem.Soc.C	1966-1971
Dalton Transactions	1972-2002
J.Chem.Soc., Dalton Transactions	2003-
J.Chem.Soc., Perkin Transactions 1	1972-2002
J.Chem.Soc., Perkin Transactions 2	1972-2002
Org.Biomol.Chem.	2003-
Proc.Chem.Soc.,London	1957-1964
Chem.Comm.	1965-1968
J.Chem.Soc.D	1969-1971
J.Chem.Soc.,Chem.Commun.	1972-1995
Chem.Commun.	1996-
Acta Chem.Scand.	1947-1973
Acta Chem.Scand.Ser.A	1974-1988
Acta Chem.Scand.Ser.B	1974-1988
Acta Chem.Scand.	1989-1999
J.Phys.Chem.	1896-1996
J.Phys.Chem.	1997-
J.Phys.Chem.B	1997-
C.R.Acad.Sci.,Ser.IIc:Chim.	1998-2001

C.R.Chim.	2002-
Anal.Sci.	1987-
Anal.Sci.:X-ray Struct.Anal. Online	2003-
J.Mol.Struct.	1968-
J.Mol.Struct.:THEOCHEM	1981-
Z.Kristallogr.,Kristallgeom.,Kristallphys.,Kristallchem. (publication suspended 1945-1954)	1921-1930, 1954-1977
Z.Kristallogr.Mineral.Petrogr.Abt A	1930-1945
Z.Kristallogr.	1978-

24 APPENDIX C: ACKNOWLEDGMENTS

The CCDC gratefully acknowledges the following copyright works, which are used under licence in ConQuest and Vista for Windows.

bsddb © 1990-2004 Sleepycat Software.
http://www.sleepycat.com/

Software from the Independent JPEG Group © 1991-1998, Thomas G. Lane
http://www.jpeg.org/

Mesa © 1999-2001 Brian Paul
http://www.mesa3d.org/

MetaKit © 1996-1999 Jean-Claude Wippler
http://www.equi4.com/metakit/

Numerical Python © 1999, 2000, 2001. The Regents of the University of California.
http://www.numeric.scipy.org

PDF and Adobe Acrobat Reader © 1987-1999 Adobe Systems Inc.
http://www.adobe.com/

Pmw © 1997-2001 Telstra Corporation Limited, Australia
© 2000-2002 Really Good Software Pty Ltd, Australia
http://pmw.sourceforge.net/

pybsddb © 1999-2001, Digital Creations, Fredericksburg, VA, USA and Andrew Kuchling.
http://pybsddb.sourceforge.net/

PyOpenGL © 1997-1998, 2000-2001 by James Hugunin, Cambridge MA, USA, Thomas Schwaller, Munich, Germany and David Ascher, San Francisco CA, USA
http://pyopengl.sourceforge.net/

Python 2.2.1 © 2001, 2002 Python Software Foundation
http://www.python.org/

The Python Imaging Library (PIL)
© 1997-2002 by Secret Labs AB © 1995-2002 by Fredrik Lundh
http://www.pythonware.com/

ReportLab © 2000-2001 ReportLab Inc.
http://www.reportlab.com/

Tcl/Tk © Regents of the University of California, Sun Microsystems, Inc., Scriptics Corporation, and other parties
http://www.scriptics.com/software/tcltk/

zlib © 1995-2002 Jean-loup Gailly and Mark Adler
http://www.gzip.org/zlib/

Allegro © 1998-2004 Shawn Hargreaves (plus contributors)
http://alleg.sourceforge.net

AllegroFont © 2001, 2002 Delerium Software
http://nekros.freeshell.org/delirium/alfont.php

FreeType2 © The FreeType Project
http://www.freetype.org

MinGW © 2004
http://www.mingw.org

Bitstream Vera Mono Font © 2004 Bitstream Inc.
http://www.bitstream.com/

py2exe © 2000, 2001, 2002, 2003, 2004 Thomas Heller
http://www.py2exe.org/